BNWL-1986
UC-32

Proceedings of the First ERDA Statistical Symposium

Edited by
Wesley L. Nicholson
Judith L. Harris

March 1976

Prepared for the U.S. Energy Research and Development Administration under Contract E(45-1):1830

Battelle
Pacific Northwest Laboratories

BNWL-1986

DISCLAIMER

NOTICE

PACIFIC NORTHWEST LABORATORY
operated by
BATTELLE
for the
U.S. ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION
Under Contract E(45-1)-1830

Printed in the United States of America
Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22151
Price: Printed Copy $8.75; Microfiche $2.25

BNWL-1986
UC-32

PROCEEDINGS OF THE FIRST ERDA STATISTICAL SYMPOSIUM

Held at
Los Alamos Scientific Laboratory
Los Alamos, NM
November 3-5, 1975

Edited by
Wesley L. Nicholson
Judith L. Harris

March 1976

BATTELLE
Pacific Northwest Laboratories
Richland, WA

PREFACE

The First ERDA Statistical Symposium was organized to provide a means for communication among ERDA statisticians. Many of us associated with ERDA and, formerly, the AEC have been concerned about the role that statistics plays in research programs. Interlaboratory communication among statisticians was infrequent and on an individual basis. A list of professional statisticians within ERDA did not exist. We often faced new problems with the feeling that elsewhere in ERDA someone had already considered and possibly solved them.

The symposium concept originated with a discussion at the 1974 Gordon Research Conference among Donald Gardiner, Keith Ziegler and myself. We learned from Milton Rose of ERDA that people at Los Alamos Scientific Laboratory had similar thoughts and had requested funding. Dr. Rose suggested we join forces and organize a meeting of all ERDA statisticians. Donald Gardiner, Ronald Lohrding, Raymond Waller and I met and decided on a program consisting of research papers and discussions of research problems with John Tukey as the keynote speaker. We included the Sandia Laboratory along with Los Alamos, Oak Ridge and Pacific Northwest Laboratories. We thought that four laboratories would provide reasonably broad participation in a program that could be put together in several months. It was expediency and not exclusiveness that limited the organizing activity and formal participation.

The organizing committee wanted the symposium to be as useful as possible. First, we wanted ERDA statisticians to get acquainted with their peers in other laboratories, their various working environments, programs and professional interests. Thus we chose to hold the symposium at an ERDA laboratory.

Second, we wanted the symposium to promote wider use of statistics on ERDA programs. Therefore, presentations of specific problems followed by discussion periods were included. Problem selection was based on ERDA goals, broadness of the applicability throughout the laboratories, and the need for statistical content for adequate solution.

Third, to learn about the statistical research going on within the ERDA community, we included research paper sessions. The after dinner talk by Dr. Jerome Friedman of the Stanford Linear Accelerator Center on graphical display and analysis of multi-dimensional data was in the framework of the research papers, since this was joint work with John Tukey under ERDA sponsorship.

Fourth, we wished to preserve the symposium as a statistical entity for evaluation by our peers; hence, the publication of these proceedings.

The organizing committee selected Oak Ridge National Laboratory for a 1976 symposium and Pacific Northwest Laboratory for a 1977 symposium. PNL is responsible for editing and publishing the proceedings of the first symposium, LASL the second, and ORNL the third.

I would like personally to thank Ray Waller, the symposium chairman and the one responsible for local arrangements, and the entire Los Alamos staff. They deserve the greatest share of credit for the success of the first symposium. Their detailed attention to all the logistics and their congeniality and hospitality made this a memorable occasion.

My coeditor of the Proceedings, Judy Harris, worked with authors on preparation of manuscripts and prepared and edited the written discussion of problems from taped transcripts. Without her dedicated work the Proceedings would not be a reality.

Finally, I would like to thank all of the participants. They accepted the challenge of honest, meaningful interaction, which assured the success of the symposium.

Wesley L. Nicholson, Editor

CONTENTS

INVITED PAPERS

USABLE RESISTANT/ROBUST TECHNIQUES OF ANALYSIS

John W. Tukey
Princeton University and Bell Laboratories
Princeton, New Jersey

This account sets out the most useful resistant/robust techniques, as we presently believe them to be, and gives indications both of what we know about them and of how they may well be used. It does not attempt to review the theory. (For moderately up-to-date background on a large slice of the field see Andrews, et al. 1972. Hampel, 1973 and 1974. Huber, 1972 and 1973.)

CONCEPTS

A function of data is "resistant" if changing a small fraction of the data, possibly arbitrarily greatly, will not change the functional value greatly. Most analytically simple functions, like $y = \Sigma x_i$, $y = \Sigma x_i^2$, $y = \Sigma x_i^2/\Sigma x_i$, for example, are not resistant. In these examples, as for the arithmetic mean, changing even one x enough can make an arbitrarily large change in y. Other functions, such as the result of "arrange the x_i according to $\leq$ and take y as the middle value," which defines the median, are quite resistant, as it is easy to see.

The term "robust" has been used by statisticians to describe various types of behavior of a function of values having a probability distribution. Most often, perhaps, the probability distribution is that of a sample from a partly known or unknown distribution of individual values (parent population). In all cases,"robust" signifies good behavior in less than ideal circumstances.

The first usage, which will not concern us here, was for significance tests or confidence intervals relating a sample to a parameter of the population from which it came. What we would now call "robustness of validity" means that the 5%, 1%, 95%, 99%, or the like associated with the test or interval applies, either exactly or to a good approximation, when the conditions are less than ideal--or, sometimes, that any major deviations are in the "conservative" direction. That this is not enough

is shown by a simple example: given a sample of 1000, draw a subsample of 10 from it at random and apply the sign test to the subsample to generate a 97.85% confidence interval for the median of the parent population. The 97.85% is exact for any parent population, but the interval uses no more than one-hundreth of the information available.

If we define the efficiency of an estimate in a given situation as

$$\frac{\text{variance of "best estimate" in that situation}}{\text{variance of this estimate in that situation}}$$

or by the result of replacing "variance" by a more appropriate concept when advisable or necessary (as, for urgent example, when no estimate has a finite variance), we will obtain a measure of estimate performance that

- is bounded by unity (by 100%) in each situation considered.
- depends, weakly or strongly, upon the situation.

An estimate is "robust of efficiency" if its efficiency is HIGH in an unusual variety of situations, whatever high may be.

We shall, in this account, say just "robust of efficiency" when the efficiencies are not necessarily very high but are nothing to be ashamed about. Using the sample median to estimate the population median is a natural example. We shall say "highly robust of efficiency" or "very robust of efficiency" when the efficiency is very high in the situations considered. An example will be given shortly.

To date, all procedures very robust of efficiency--and, to a lesser degree, procedures only (but not very) robust of efficiency--are resistant. So far as we can see, this is inevitable. Hence the splice-word--resistant/robust (or robust/resistant).

GOOD PRACTICE

What are today's obligations of good statistical practice? I suggest they include these two:

- where one or more analyses robust of efficiency are available, one (or more) of them should be among those applied (in parallel) to the data, and its (their) results should be compared with those of the other analyses used.

- where one or more analyses highly robust of efficiency are available, and we are concerned to extract all the information from the data, one (or more) of them should be among those applied to the data, and little credence should be given, in the absence of special considerations, to the result of any analysis that does not agree reasonably well with the result of the resistant and highly robust analysis (analyses).

This means that any analysis based upon:

- arithmetic means
- moments
- least squares

to name a few standard cases, needs to be at least accompanied by a resistant/robust analysis if an appropriate one can be found. Some may think this position goes a long way, but experience--with situations where conventional analyses have gone awry without giving warning--suggests that it is essential to go at least this far.

Let us then turn to specific kinds of analysis and summarize where we stand. The different problems to be considered, which together represent most of the foundations of statistical analysis, fall into two broad classes:

- problems where estimated quantities do not affect the identification of unusual residuals (these include the naturally simpler processes, like location and univariate scaling).
- problems where the identification of unusual residuals may depend critically on the resistance of the estimate of the fit (these include many important standard problems, including regression, smoothing, correlation estimation and estimation of vector dispersion).

We shall take our problems up in this order. Thus an important emphasis on resistance/robustness for residuals will be delayed to Part II.

PART I

WHERE FITTING IS NOT NEEDED TO IDENTIFY POTENTIAL "WILD SHOTS"

The main problems here are those of scalar (single or multiple) centering or widthing. In such scalar cases the potential "wild shots" are clearly the lowest and highest values. We know which points to examine most carefully.

The case of vector centering, which we take up at the close of this part, is perhaps a transitional case. Actually, of course, we do need vector centering to identify vector "wild shots", but we need vector dispersion even more.

INDIVIDUAL CENTERING

SYMMETRIC CENTERING

Given values $y_1, y_2, \ldots, y_n$ which we have agreed to analyze as if they were a sample from a symmetric distribution--or were generated symmetrically in a somewhat more complicated way--how should we estimate the center about which the underlying distribution--or underlying process--is symmetrical?

Not so long ago, our answer would have been easy and dangerous: take the arithmetic mean.

At the height of the nonparametric era we would have been much safer, though somewhat wasteful--we would have taken the median.

Today we would like to be safe without being sorry.

TRIEFFICIENCY

We have long looked to efficiency, to the ratio

$$\frac{\text{smallest variance deliverable}}{\text{variance actually delivered}}$$

as a criterion of estimate quality. Clearly the answer has to depend on the situation we are considering, Gaussian samples and Cauchy samples are unlikely to give the same number. Two remarks and a warning are now badly needed.

First remark: "deliverable" often has to refer to the best we know how to do--in which case we might (and will try to) refer to "apparent efficiency"--or to a theoretical bound (which we will here avoid). This is because we do not know exactly what the smallest variance deliverable is in most situations.

Second remark: "variance" has sometimes to be replaced by a related measure (as noted above).

Warning: knowing the efficiency--or the apparent efficiency--in just one situation is almost never enough. We need, at least, to explore extreme situations that will stress our estimates in each of the recognized ways we have found to be particularly threatening.

In the case of symmetric centering--say for batches of 20 values, the case for which the most extensive calculations have been made,--we today recognize three kinds of threats. Thus we need to consider at least three situations. To challenge estimating procedures as severely as we can, we would like the challenging situations to be rather unrealistically extreme each in its own way.

The three situations for which most computation has been done, which do seem about equally unrealistic, are these:

- the pure Gaussian distribution (too nice).
- a situation combining a sample of 19 from one Gaussian with one observation from another Gaussian with the same center and 10 times the spread (too black-and-white in its non-constant variance).
- the slash distribution, which is representable as the ratio of a unit Gaussian to an independent unit rectangular (on [0, 1]), which avoids the unrealistic central peakedness of the Cauchy, yet whose tails are of Cauchy type (too stretched-tailed).

If a centering procedure works well in all these three situations, we can hope it works even better for all the more realistic situations that lie "between" them. Thus it is natural to define a "triefficiency" by:

triefficiency = the least of these three efficiencies

and to seek for a centerer that has high triefficiency.

SOME RESULTS--THE BIWEIGHT

The classical centerers are, by now, the (arithmetic) mean, the median, and the trimmed means (in which the j highest and j lowest values are set aside, and the arithmetic mean of the remainder provides the estimated center). For batches of 20 observations, the Princeton Robustness Study (Andrews, et al, 1971) offers us apparent efficiences at the three corners for these classical estimates, namely:

	Pure Gauss	One Wild	Slash	Triefficiency
Mean	100%	17%	0%	0%
Trimmed 1 + 1	98%	92%	16%	16%
Trimmed 2 + 2	95%	94%	41%	41%
Trimmed 3 + 3	94%	94%	60%	60%
Trimmed 5 + 5	84%	87%	86%	84%
Median (10 + 10)	67%	72%	87%	67%
(Now known to be attainable)				(92 - 93%)

This is a background against which we might seek a centerer that is highly robust of efficiency.

Some of the simplest high-performance estimates that we know are of the form

$$\hat{\text{center}} = \frac{\Sigma w_i \, y_i}{\Sigma w_i}$$

where each w_i depends on the dimensionless or scaled form of $y_i - y'$, where y' is the median of the y's, which we take as

$$u_i = \frac{y_i - y'}{cS}$$

where c is a number and S a measure of spread of the y_i, usually the MAD (median absolute deviation from the median), the H-spread, or the inter-quartile range.

The choices c = 9 and

$$w = w(u) = \begin{cases} (1-u^2)^2, & \text{for } u^2 < 1, \\ 0, & \text{else.} \end{cases}$$

give a triefficiency, in samples of 20, above 90%. (By complicating matters, we can squeeze out a little more triefficiency, but it may well not be worth it.)

Since $(1 - u^2)^2$ is sometimes called a bisquare, it is perhaps natural to call the center just described a (one-step) biweight.

Drawing on the general results of the Princeton Robustness Study, and the general insights it produced, we can be quite sure that the biweight does quite well indeed for batches of 10 or more y's--and can be reasonably sure of its performance down to, or close to five.

Tony Quon, in an almost completed Princeton Ph.D. Thesis, has shown that there is little to gain over using the median for batches of three. We can expect a similar result for n = 4. Thus a good rule runs as follows:

Number of y's	Centerer to be used
up to 4	median
5 or 6	take your choice (biweight favored)
7 or more	biweight.

(It may be of interest that about 10^3 distinctive estimates have been tried out, as well as about $10^{4.5}$ 50-50 linear combinations of the same. Only a new idea is likely to raise the triefficiency for n = 20 appreciably above 92-93%.)

ESTIMATING VARIABILITY

The variably-weighted mean used in the (one-step) biweight can be regarded as the first step in an iteration

$$T_{k+1} = \frac{\Sigma w_i^{(k)} y_i}{\Sigma w_i^{(k)}}$$

where the $w_i^{(k)}$ are calculated from the $(y_i - T_k)/cS_k$. Like the Newton-Raphson iteration

$$T_{k+1} = T_k + \frac{\Sigma\psi(\frac{y_i - T_k}{cS})}{\Sigma\psi'(\frac{y_i - T_k}{cS_k})}$$

where $\psi(y) \equiv u \cdot w(u)$, this iteration converges to a solution of

$$\Sigma\psi(\frac{y_i - T}{cS}) = 0$$

whose asymptotic variance is

$$\frac{\Sigma\psi^2}{[\Sigma\psi']^2}$$

or, as is asymptotically equivalent,

$$s_*^2 = \frac{\Sigma\psi^2}{[\Sigma\psi'][-1+\Sigma\psi']}$$

which reduces, for $\psi(u) \equiv u$, to

$$\frac{\Sigma(y-T)^2}{n(-1+n)}$$

the classical variance of $\bar{y}$ which is the solution of

$$\Sigma\psi(y-T) = 0$$

for this $\psi(\cdot)$.

INTERVAL CENTERING

Alan Gross (1973, 1976a) has studied the interval estimation of the center of a symmetric distribution, finding that, in our present notation *, the interval

$$(\text{biweight} - t_{.7\nu}\, s_*,\ \text{biweight} + t_{.7\nu}\, s_*]$$

*The inclusion of a -1 in the denominator of s_*^2 raises the degrees of freedom at least from $.5\nu$ to $.7\nu$. The inclusion of -2, instead, which may prove helpful, might raise the degrees of freedom to ν.

where $t_{.7\nu}$ is the 5% value of t on seven-tenths* the usual number of degrees of freedom, is quite closely a 95% confidence interval for the population center, not merely for the Gaussian case but for a variety of other situations. (He tried a variety of alternate centers and width indicators. Only those closely similar to the biweight and s_* worked nearly as well.)

His results establish a good performance for $n \geq 10$. He anticipates satisfactory performance down to n = 8. For smaller n, only nonparametric procedures like the sign test which can give, non-resistantly and thus surely not highly robustly, 98.4% confidence for n = 7, 96.8% for n = 6, 93.8% for n = 5, and 87.5% for n = 4 or, resistantly, 87.5% confidence for n = 7 and 75% for n = 6--or the one-sample Wilcoxson test, which is somewhat more flexible but has the same extreme % levels, is known to give reliable interval estimates (of unknown efficiency) for the population center.

A NON-PROBLEM?

We have discussed symmetric centering. What about unsymmetric centering? What, indeed.

We all know what we are trying to estimate in the symmetric centering case--the mean, the median and all the mids (the means of symmetric % points) coincide, making a unique target. But what is our target in the unsymmetric case?

We can take one of two routes:

- We can think about each of as many realistic special cases as we can find where a specific target makes specific sense; or
- We can try to find a measure of location--an estimate commuting with the addition of a constant--that has high robustness of efficiency, and then declare that whatever it appears to estimate is the center we wanted to estimate.

Either of these is possible. Neither is what we intuitively thought of on first meeting the words "unsymmetrical centering." (A similar problem arises from the beginning in dealing with width as will be discussed later.)

---

*See footnote on previous page.

TOTALS

MONTE CARLO--AND INSURANCE

There are situations, two of the most striking being Monte Carlo calculation and the insurance business, where our concern is really with totals. Real Monte Carlo, as distinguished from simple experimental sampling, operates by a "swindle", by replacement of the original problem by one that can be demonstrated to have the same answer. Almost always the answer is an arithmetic mean--essentially a total, since probability theory almost always deals with such. (Recall that a probability is a mean of the corresponding indicator function.) Almost always the values to be totalled are non-negative, but have a stretched-out tail to the right.

We would like to think that we are dealing with a severely asymmetric problem of centering and use a resistant-robust center. But we dare not. For the only target we are allowed to shoot at is the arithmetic mean. Any center that is resistant will be biased downward. We must take full account of occasional very large values.

NO ESCAPE

We are going to have to summarize our observation by $\bar{y}$.

For the present we have little, of anything, better to do than to use the conventional s^2 and Student's t though we need to do rather better. To do better we need, overtly or covertly, to make some allowance for both skewness and elongation of our distribution. Doing this by moments is quite unsatisfactory (we will soon comment on an alternative).

We need to remark that Student's t for a very elongated sample, for instance n-1 values close to zero and one value large, comes out close to unity. Thus confidence intervals got by Student's t (at any usual level) turn out to be conservative, not liberal. What we have to fear with Student's t, so far as validity goes, is the squeezed-tail case. Stretched-tail cases, if recognized and properly assessed, offer us an opportunity to learn how to say more than Student's t does.

G/H TECHNIQUE, PERHAPS?

Suppose z has the unit Gaussian (or unit neutral) distribution--too often (mis)called the unit normal distribution. A convenient family of skewed distributions are those of $(e^{gz}-1)/g$ for various g. A convenient family of elongated distributions are those of $z \exp(hz^2/2)$ for $h > 0$.

A little rough calculation based on these examples leads to plotting $(y_{-p}+y_p)/2$ against z_p^2, where y_p is the empirical upper p% point in our sample, y_{-p} the lower p% point, and z_p^2 the square of the one-sided unit-Gaussian p% point, with the anticipation that the slope, if we detect any, will tell us about something like "g". Similarly, a plot of $\log(y_p - y_{-p}) - \log(2z_p)$ against the same z_p^2 has a hope of telling us, by its slope, about something like "h".

It is by such techniques that there seems hope for improving Student's t in the face of skewness and elongation.

COMPARATIVE CENTERING

THE TWO-SAMPLE CASE

How are we to compare the centers indicated by $y_{11}, y_{12}, \ldots, y_{ij}$ and $y_{21}, y_{22}, \ldots, y_{2k}$ in a resistant and robust way? The analogy with the one-sample case is strong enough for us to be sure of very good performance if we use

$$y_{1B},\ s_{1*}^2,\ 0.7\nu_1,\ y_{2B},\ s_{2*}^2,\ 0.7\nu_2$$

where y_{1B} and y_{2B} are the biweights and s_{1*}^2 and s_{2*}^2 are the associated s_*^2's and $0.7\nu_1$ and $0.7\nu_2$ are the associated 0.7 of degrees-of-freedom.

The careful analysis, allowing for possible difference in width in the two situations takes

$$s_{D*}^2 = s_{1*}^2 + s_{2*}^2$$

$$\frac{1}{0.7\nu_D} = \left\{\frac{s_{1*}^2}{s_{D*}^2}\right\}^2 \frac{1}{0.7\nu_1} + \left\{\frac{s_{2*}^2}{s_{D*}^2}\right\}^2 \frac{1}{0.7\nu_2}$$

$$t = \frac{y_{1B} - y_{2B}}{s_{D*}}$$

$$\nu = .7\nu_D$$

The more off-hand analysis takes

$$s_{D*}^2 = \frac{n_1\nu_1 \, s_{1*}^2 + n_2\nu_2 \, s_{2*}^2}{\nu_1 + \nu_2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

$$t = \frac{y_{1B} - y_{2B}}{s_{D*}}$$

$$\nu = .7\nu_1 + .7\nu_2$$

and (1) will be tighter than the careful analysis, and (2) can be in trouble if the widths are not the same.

I would have no hesitation in proceeding along whichever of these routes seemed appropriate, confident that the answers would be safer and almost as precise as (therefore probably more accurate than) those obtained from sample means and sample s^2's.

IF THERE MUST BE A REFERENCE

Sometimes it is necessary to have a published reference for the technique to be used. In the two-sample case, this may, for the present, cause giving up the (more preferable) biweight in favor of (less preferable, but more referable) trimmed t (Yuen and Dixon, 1973, Yuen, 1974).

MULTIPLICITY

SIMULTANEOUS AND QUANTITATIVE: BONFERRONI

If we are to make several--or many--quantitative statements with a requirement of conventional certainty, 95% or 99%, we must make each individual statement with very much higher certainty. There can be no escape from this--since each statement is free to be wrong separately.

The basic, simple calculation is that associated with the name of Bonferroni. Given a group of K statements, each at an error rate of (p/K)%

we will have, on average, no more than K(p/K)% = p% statements wrong per group. Therefore, a fortiori, the chance of one or more statements wrong is no more than p% per group.

The need for Bonferroni implies a need for understanding the more extreme % points of our statistics--e.g. the 0.05% point when K = 100 and p = 5%. Our information here is quite limited, but what insight we have suggests that--even far out in the tails--a two-sample biweight comparison is still much safer than a two-sample arithmetic mean comparison. Here is a place where more knowledge would be particularly welcome.

BELFERRONI

If we are prepared to control the ratio

$$\frac{\text{number of batches of statements with at least one error}}{\text{number of batches of statements}}$$

instead of the ratio

$$\frac{\text{number of statements in errors}}{\text{number of batches of statements}}$$

we ought to be able to tighten the statements just a little. In the ideal case--everything Gaussian of equal variance--when arithmetic means are used we can calculate factors to supplement K, so that we use, say, p/KLM in place of p/K where L and M are less than unity. To do this requires comparing conventional % points of the Studentized maximum modulus (limits on all values) or the Studentized range (limits on all simple comparisons) with the tails of Student's t. Tables in this form are in preparation.

When available, I would be quite happy to use these tables in connection with multiple statements about values or comparisons of biweights--happier, actually, than to use them with values or comparisons of arithmetic means. (In the long run, of course, we want to know more about just how well they apply. Our uncertainties here, however, must be small compared to our uncertainties about the far-tails of biweight statistics.)

ONE-HUMP PHILOSOPHY

The case of qualitative statements about all simple comparisons is a special one. One philosophy that has, more or less tacitly, been building in recent decades emphasizes the situation where the population centers are

themselves distributed in a single-humped fashion. The case where population centers are samples from a single-humped (perhaps Gaussian) distribution seems likely to play a special role in assigning nominal levels of significance.

There seems no reason to expect difficulty in taking over to biweights whatever approaches we settle upon for arithmetic means.

SPLITTING AND MULTI-HUMP

Sometimes what is needed is only to split up the values to be compared as well as we can, to do as good a qualitative job as we can. Here unpublished work of Roy Welsch will soon provide us with a well understood and sharp tool, at least where arithmetic means are concerned. The writer would have no hesitation using Welsch's tables on biweight estimates and their associated s_*^2.

A little more concern could arise if either different s^2 or different s_*^2 are associated with different ones of the values to be compared. But it should be much better to use Welsch's tables than to wait.

Welsch's techniques should also serve, in any of these four cases, to decompose multi-humped situations into pieces adequately treatable by one-humped techniques.

WIDTH

FREE OR BOUND

We may ask of a measure of width--of scalar dispersion--only that it:

- indicate relative width as between situations differing only by scale change.
- do so sensitively, perhaps as judged--separately in each of many situations--by the efficiency of its logarithm.

This is free widthing.

Not all of us have thought deeply enough about the problem of estimating width to recognize how often we do not know what we want to estimate. In part, this is because the books concentrate on the Gaussian case, where there are fixed ratios among, for instance:

- the population standard deviation.
- the population mean deviation.
- the interquartile spread.
- the various other spreads between p% points for different p.
- the limiting case as p% → 50%, namely the reciprocal of central density.

If we consider a variety of distribution shapes, the interconversion factors among such measures of population spread differ widely from one shape to another. Yet an estimate that estimates any of them is clearly an estimate of width.

There are situations where this freedom is removed. If we are assessing the stability of a simple arithmetic mean, we need to estimate the population standard deviation. If we are assessing the stability of a simple median, we need to estimate a p% spread, for p suitably close to 50. There are, then, many cases where we want to add to the two requirements with which we started this section a third, namely:

- what is estimated in each of many situations tracks, at least bearably well, some target value (perhaps the square root of the variance of some other computed quantity), as the population shape changes.

When we add this third requirement, we have a bound widthing problem.

NOSE IN THE BUTTER

We now have, in work at Princeton under ARO auspices, some very useful results on free widthing. We do not yet have the degree of completeness that exists for symmetrical centering, but we have come a long way--and to a most pleasant result.

If we take the same sample size and three corners that have guided much of the work on symmetrical centering:

- first, sample of 20 from the pure Gaussian,
- second, samples of 19 from Gau (0,1) combined with samples of 1 from Gau (0,100),
- third, samples of 20 from the slash distribution, with its Cauchy-like tails,

we are now able to provide an apparent efficiency of estimate of log width of at least 86% (Lax, 1975) at all three corners. The best that conventional estimates of width seem to do is 35% for the MAD--the median absolute deviation (from the median).

What delivers this great improvement? None other than our friend Alan Gross's denominator--the natural asymptotic estimate of the biweight's sampling standard deviation s_*. As the Danes say, we have "fallen with our nose in the butter". For the present, at least, OUR BEST FREE WIDTHER is ALSO the BEST WIDTHER BOUND to our BEST CENTERER. How could things be simpler?

While other estimates of width may prove somewhat better, we are now quite certain of a permanent place for a widther that is both:

- excellent as a free widther and
- bound to the variability of an excellent centerer.

The advantages of such a combination, something that does for a variety of situations what $\bar{y}$ and s do for the over-utopian Gaussian case, are just too great for us not to seize upon them--hard.

SOME OPEN AREAS

All our results are for sample size 20. In view of Gross's results, it seems safe to use them for guidance for samples $\geq$ 10. What to do for smaller samples is not yet clear, nor is it clear to what extent building down to smaller samples may help us with the interval centering problem for $n < 10$.

The only estimate of variability for s_* so far at hand is that offered by jackknifing, where jackknifing the cube-root or log of s_* seems natural.

As written down, s_* is, roughly, $(1/\sqrt{n})$th of an estimate of width--s_*^2, after all, estimates the variance of an efficient summary. So long as we deal either (a) with only one sample size or (b) with only one use for width--assessing the stability of an efficient centerer--we need have no concern, But otherwise we need to know just what to multiply by--is it $\sqrt{n}$ or $\sqrt{\Sigma\psi'}$ or what?

Despite all these questions, a rather large-scale use of s_* seems inevitable.

VECTOR CENTERING

DISTINCTIVE OR AFFINE

The problem of resistant/robust centering of vectors, like so many problems in the analysis of vectors, comes in two extremes:

- distinctive centering, where we are to work in a prespecified coordinate system.
- affine centering, where we are to work in such a manner that the answer reached is independent of the coordinate system.

There is some ground for suspicion that some intermediate formulation will come to be of real importance, but this has not yet happened.

TODAY'S STATUS

Distinctive point estimation of vector centers requires only point estimation of centers for each coordinate, something we can do with high robustness of efficiency.

Distinctive interval estimation of vector centers by parallelopipeds parallel to the distinctive coordinate planes can be done with equal ease, using Gross's techniques on each coordinate, but parallelopipedal confidence intervals are often quite wasteful.

Distinctive interval estimation of vector centers by ellipsoids requires corresponding techniques for assessing vector dispersion. As discussed below, resistant/robust techniques are probably well on their way rather than being ready on the shelf.

Affine point estimation of vector centers in a resistant/robust way seems also to require resistant/robust assessment of vector disperion.

Affine interval estimation of vector centers in a resistant/robust way seems to require resistant/robust assessment of vector disperion, twice over.

AN EXAMPLE

As an example that resistant vector centering problems need not be trivial, look at Figure 1, taken from the study reported in Chen, Gnanadesikan, and Kettenring (1974). This plot shows the first two principal components for 14 economic characteristics of 29 chemical companies. (No

errors, only a maverick chemical company, one that is not noticeable in any one of the 14 variables alone.) Clearly the inclusion or suppression of the company plotted in the lower right corner alters substantially both centering and dispersion for this batch of data.

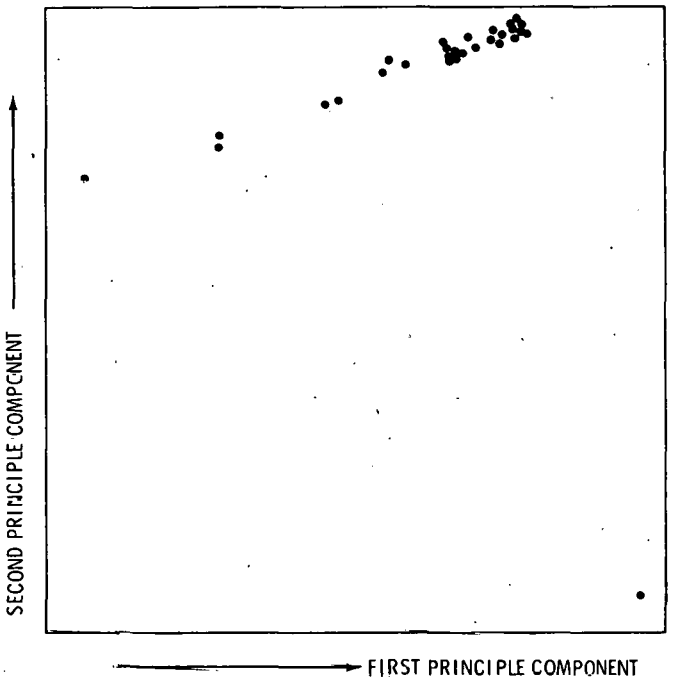


FIGURE 1. First Principal Components of Economic Characteristics of 29 Chemical Companies

PART II

PROBLEMS WHERE FITTING IS ESSENTIAL IN IDENTIFYING "WILD SHOTS"

In all these problems, resistant techniques seem essential, if a good job of identifying "wild shots" is to be possible. In a two-way analysis, for example, a few "wild shots" can so distort an analysis by means as to easily conceal their own presence, as well as providing a misleading fit. Thus resistant methods are essential if we want to:

- identify likely wild shots, so that their cause may be inquired into,
- identify likely wild shots, so their setting aside will let us see the actual general pattern,

- make a fit that reveals the desired general pattern.

It is essential to have resistant/robust techniques. There is no substitute.

The discussion of this part falls naturally into four subparts:

- regression
- smoothing
- patterned fitting
- correlation and dispersion

INDIVIDUAL REGRESSION

ITERATED BIWEIGHTS

A number of authors (e.g. Andrews, 1974, Beaton and Tukey, 1974, Hill and Holland, 1974, Hoaglin, Holland and Welsch, 1975) have made use of iterative weighting in regression situations. The choice of just which iteration was used has varied somewhat, but there is no reason to believe that this choice is particularly important. So we will here recommend the use of a biweight.

The iterative computation then involves:

$$u_i = \frac{y_i - \hat{y}_{old}}{cS_{old}}$$

where S_{old} summarizes, perhaps as the MAD, the sizes of all the $y-y_{old}$, and

$$w = \begin{cases} (1-u^2)^2, & u^2 \leq 1, \\ 0, & u^2 \geq 1, \end{cases} \qquad \text{(subscript i suppressed)}$$

The y_{new} arises by weighted regression, in which the ith data set receives weight w_i. (If weights W_i were inherent in the regression problem, we may want either just to use weights $w_i W_i$ or to replace $y-\hat{y}_{old}$ by $\sqrt{W}(y-\hat{y}_{old})$ in the calculation of the fitting weights, w_i.)

How many iterations are likely to be needed must, among other things, depend upon how many constants are being fitted. We know very little about this. So we plan to iterate to reasonable stability, which does not usually require many iterations.

Where we start is important, mainly as a computational matter. If we start from an unweighted least-square fit, we are likely to need more iterations. If we start from, say, a least-absolute deviation fit, we will have had to expend more computation to get to our start.

THE ANDREWS CASE

The sort of situation illustrated in Figure 2 has been forcefully called to our attention by David Andrews. We know of no escape from the need of two analyses, one corresponding to each of the lines in the picture. It is then a matter of wisdom, subject-matter knowledge, and inquests over the bodies of individual exotic data-sets to decide how to go further and how to report the result.

This problem is not restricted to the simple case just illustrated. Turn back to Figure 1, and assume what is shown are two x-variables. However y behaves, the fit for:

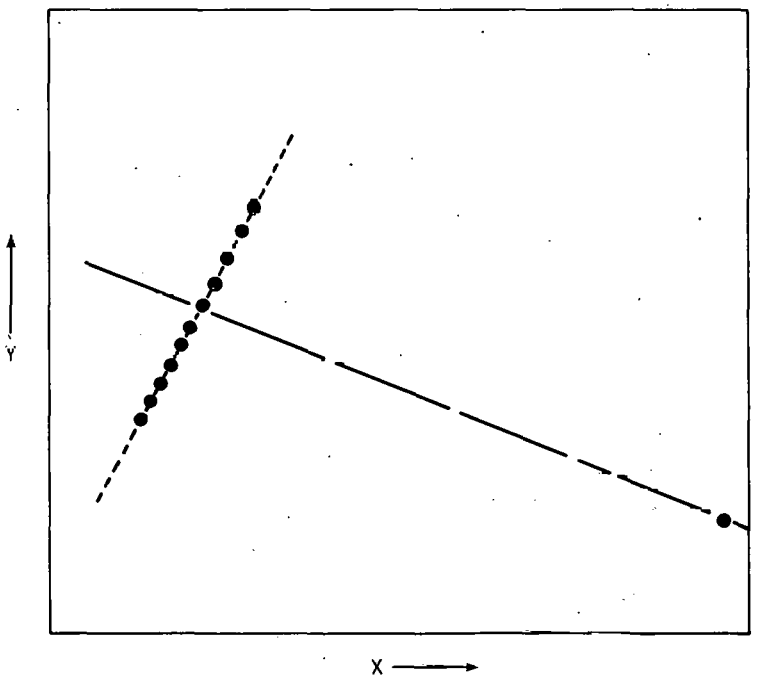


FIGURE 2. An Andrews Configuration: Some Data Points With Two Fits (either line may be relevant)

- all data points
- all data points except the exotic one

are certain to be quite different. (Even if the fits resemble one another, the uncertainties assigned to the fitted coefficiencts will not.) With several x's even the recognition that we face such a case is not easy. (And the work of N.E. Day, 1969, suggests that rather drastic situations may not be uncommon when the x-vectors are samples from a Gaussian distribution of vectors.)

A variety of suggestions have been made (e.g. by Andrews and by Mallows) of ways in which we might be reasonably sure to get the fit setting aside exotic points. To date there seems to be no procedure which guarantees us finding all the plausible solutions.

Using a variety of random starts for the biweight iteration may help us; but when we do this, we are unlikely to ever know how many other solutions we missed.

WHAT VARIANCE?

Once we have our robust/resistant regression, which we expect to be better determined than the least squares regression, how do we judge its accuracy? A simple approach is to forget that the weights are relative and calculate a variance-covariance matrix as if the weights were fixed.

A variety of more complicated formulas have been proposed for consideration. Work at National Bureau of Economic Research-Cambridge (Hoaglin, Holland, Welsch, et al) has thrown some light on this question. Recent work by Gross (1976b) on intervals for linear combinations of a and b, where a + bx is being fitted, seems to indicate that different "variance" estimates require different "t" values, but otherwise perform similarly.

If this proves to be so in more general fits, and for "F" as well as for "t", we will be indeed fortunate.

In the meantime, using the estimated variances and covariances neglecting the iterative nature of the weighting seem a useful guide to the stability of our results. (Unpublished results of Mallows would suggest expanding variances and covariances by a factor of $(1.07)^2$ when using biweights with c=9.)

AN AFFINE OBLIGATION?

When we fit

$$b_1 x_1 + b_2 x_2 t + \ldots + b_k x_k$$

we may consider $(b_1, b_2, \ldots, b_k)$ as a vector. If we do, the carriers x_1, x_2, ..., x_k determine--mainly by their omission--a set of coordinates in the space of vectors. Replacement, say, of x_1 by

$$x_1 - .05\, x_2 + 13 x_3 + \ldots + .09 x_k$$

introduces a new set of coordinates and represents identically the same fit--the same abstract coordinate vector--in a new set of coordinates (one where b_2, b_3, ..., b_k have all been altered). We could ask for the estimation of the b-vector to be affine invariant.

Iterative reweighting from an affine invariant start will give an affine invariant result. So the question of how hard should we work to obtain an affine invariant result reduces to--

- should we insist on an affine invariant start?
- should we insist on an affine invariant assessment of vector dispersion for the regression coefficients?

These questions are not as simple as they seem.

Suppose we start from (0,0, ..., 0) as the initial set of b's. Is this an affine invariant start? In a narrow sense--invariance under all HOMOGENEOUS non-singular linear transformations--yes. In a broader sense--invariance under all non-singular linear transformations, homogeneous or not--no. Which sense is appropriate? Can any fit be affine invariant in the broader sense?

This whole question is clearly presented as a teaser.

OTHER REGRESSION

COMPARATIVE REGRESSION

There seems to have been nothing explicit to report here. We presumably add together whatever variance-covariance matrices are judged appropriate for the individual regressions.

AUTOREGRESSION

Work by Douglas Martin and Lorraine Denby, not yet submitted for publication, indicates that ψ-w techniques, including iterated biweights, do well here also.

SMOOTHING

BASIC IDEAS

The idea of smoothing a sequence is old; classical performance is at best poor. There are two distinct reasons why arithmetic means, simple or repeated, do not perform well.

The first failure is shown by a clear tendency to pull hills down--and valleys up--as well as reducing ripples. Repeating the smoothing--resmoothing--as in

$$z_i = \frac{y_{i-2}+y_{i-1}+y_i+y_{i+1}+y_{i+2}}{5}$$

followed by

$$v_i = \frac{z_{i-2}+z_{i-1}+z_i+z_{i+1}+z_{i+2}}{5}$$

or

$$v_i = \frac{z_{i-1}+z_i+z_{i+1}}{3}$$

makes this deficiency worse.

The cure is easy. Rerough instead of resmooth!

Having done a first smooth, perhaps

$$z_i = \frac{y_{i-2}+y_{i-1}+y_i+y_{i+1}+y_{i+2}}{5}$$

form also the "rough", satisfying given data $\equiv$ smooth PLUS rough, namely

$$q_i = y_i - z_i$$

and SMOOTH THIS ROUGH, finding, perhaps

$$t_i = \frac{q_{i-2}+q_{i-1}+q_i+q_{i+1}+q_{i+2}}{5}$$

or

$$t_i = \frac{q_{i-1}+q_i+q_{i+1}}{3}$$

AND THEN ADD TOGETHER the TWO SMOOTHS, find $z_i + t_i$ as one's smoothed values. (When, as in our examples above, each smoother reproduces every straight line identically, the combination will reproduce every cubic identically.)

Second, as something involving means, smoothing by running means is far from resistant. One horribly wild point, for example, is smoothed out into several not quite so horribly wild points. This is often unbearable.

Again a cure is simple. As a first step, replace running means by running medians. This will supply the resistance, but need not make things smooth enough. (Monotone sequences, for example, are unaffected by smoothing with running medians.)

3RSSH, TWICE

An elementary procedure that is easy to learn, and produces better smooths than you are likely to begin by hoping for, makes use of these three kinds of components:

- running medians of 3
- the end value rule

  z_0 = median $(3z_1-2z_2, y_0, z_1)$

  where z_1 and z_2 are already obtained smoothed values.
- hanning, forming

  $$\frac{z_{i-1}+2z_i+z_{i+1}}{4}$$

  by one algorithm or another.

The rules of procedure are as follows:

- repeat the use of running medians of three--so long as this changes the sequence. During this phase the end values are "copied on"

until running medians of three have no further effect, at which point the end value rule is applied to each end. (All this is called 3R, for "three, repeatedly".)

- identify every flat-topped hill or flat-topped valley of width exactly two; and act as if the sequence had been split in the middle of each of these two-flats in turn, applying the end-value rules to each new end thus formed.
- after this has been done for each two-flat in turn, repeat 3R. (These two steps are called "S", for splitting.)
- repeat these two steps once more (another S).
- apply hanning, copying on end values (called H).
- find the rough, and repeat the whole process on the rough, adding the resulting smooth to the original smooth. (When "reroughing" is by a copy of the original smoother, we speak of "twicing".)

For examples, see Tukey (1976).

The resulting procedure, called, for its components, "3RSSH, twice" is resistant and moderately effective.

HIGHLY EFFECTIVE SMOOTHERS

Whether or not it is possible, what would be a utopian dream for an ideal resistant smoother? On the one hand, it would chop off isolated "wild shots", leaving hardly a trace. On another, it would follow every smooth sequence very, very closely.

Clearly no linear smoother can come close to this, since its behavior for any sequence has to be the sum of its behaviors for the sinusoidal components of that sequence and since a wild shot's sinusoidal components involve slow (smooth) frequencies as well as high ones.

Exactly how close nonlinear smoothers can come to meeting this dream we do not know. But Paul Velleman (1975) has shown that one can smooth long sequences so that

- for exactly sinusoidal input, whose output must be a multiple of that sinusoid plus multiples of its harmonics, no harmonic is less than 25 db (a factor of 18) smaller than the input.

- the multiple of each sinusoid behaves the way we would hope: close to 1 at low frequency, close to 0 at high frequency, smooth transition between.
- stretched-tail or spiky noise is adequately smoothed away.

While not reaching the impossible ideal dream, this goes farther toward it than many would have deemed possible.

His techniques combine biweighting with cosine-arch coefficients and reroughing.

ONE APPLICATION

Resistant smoothers work quite well on most unlikely material. If we have a scatter plot of (x,y) pairs, we can:

- sort on x
- eliminate ties by replacing all pairs with the same x by their median (if few) or biweight (if several or many)
- smooth the resulting sequence of y's
- smooth the corresponding sequence of x's
- plot the broken line through the resulting (x-smoothed, y-smoothed) pairs.

What we get from a respectable (hence resistant) smoother, "3RSSH twice", for example, is often very illuminating and very helpful.

Scatter plots inevitably direct our attention to the outside points, not to a good general picture of what is happening.

A scatter plot unsupplemented by one or more resistant smooths is at best a lost opportunity. At worst it can be a disaster.

"PLUS" ANALYSES

BY MEDIANS

Fitting $a_i + b_j$ to a table of values of y_{ij} is a very widely used process. For complete rectangular arrays (y_{ij}) we used to describe it noniteratively in terms of row means and column means.

If we want to describe fitting in terms of medians, which we often find very useful, it is essential to describe the fitting iteratively.

The simplest description starts with one set of stripes, either rows or columns, and proceeds as follows:

- find medians of one set of stripes (the A-stripes), including one stripe of fits (initially probably zero).
- subtract these from the data (the residuals) and add them to the fit.
- find medians of the other set of stripes (the B-stripes), including one stripe of fits.
- subtract them from the residuals and add them to the fit.
- repeat, perhaps for four find-and-subtract steps, perhaps to apparent convergence.

By now there is quite extensive experience with this process, which does very well in being resistant and reasonably robust.

Since it uses medians, we cannot expect it to show high robustness of efficiency, unless there are no more than, say four rows and four columns. (Or perhaps no more than six?)

BY BIWEIGHTS

We can get high robustness of efficiency for larger numbers of rows or columns, we are quite sure, by replacing the medians in the iteration process by biweights. I know of no explicit numerical studies, but I have no appreciable doubt--this is a high quality process.

GENERALIZATIONS

Extensions to fitting y_{ijk} with $a_i + b_j + c_k$ or with $a_i + b_j + e_k + d_{ij} + e_{ik} + f_{ik}$ are relatively easy, with the latter (surprise!) the easier. To perform the latter analysis, we can define three kinds of fibers:

- all data points with i and j fixed.
- all data points with i and k fixed.
- all data points with j and k fixed.

and then iterate cyclically through transfers from residuals to fit (and from fit to fit) corresponding to A-fibers, B-fibers, and C-fibers. Generalization to any unrestricted factorial pattern is easy.

Restricted factorials can be handled without difficulty by using (1) an unpublished algorithm of P. J. Claringbold to determine which kinds of

fit are desired and (2) insertion of missing value indicators to indicate which fibers are to be omitted from the find-and-transfer process.

ANALYSIS OF VARIANCE--A PROSPECTUS

By analysis of variance, as distinct from PLUS analysis, I mean the ideas and techniques that revolve around the classical analysis-of-variance table and its improvements. These ideas are concerned, on the one hand, with the breakdown of the variance of an observation into portions (inadequately mean squares, more adequately variance components) "assignable" to different sources, and, on the other and more frequent hand, with the use of one mean square--or linear combination of mean squares--as an "error term" for drawing conclusions--either of significance or confidence, about some or all of the values appearing in some effect or interaction. In the latter case, the classical F test has been replaced, in many cases, by multiple comparison procedures (compare with multiplicity section).

In this area, we can see the dawn, but not quite the sun. The serendipitous appearance of s_* as both an indicator of the variance of biweights and a high-quality estimate of width suggests quite clearly that each of our mean squares will need to be replaced by two numbers, one of which is an s_*^2, to be used as--or as part of--an error term. The s_*^2's, as well as effects/interactions will be based on a resistant/robust analysis of the sort just described.

Details soon, perhaps.

INDIVIDUAL CORRELATIONS--QUARTER SQUARES

Once upon a time, multiplication was done by squaring, using the identity:

$$uv \equiv \left(\frac{u+v}{2}\right)^2 - \left(\frac{u-v}{2}\right)^2$$

and Tables of Quarter Squares. Today the existence of a good resistant/robust widthing procedure implies that of a plausible resistant/robust correlation calculation, based upon

$$u = \frac{y}{\text{width of } y}$$

$$v = \frac{x}{\text{width of } x}$$

and either

$$\text{"correlation"} = (\text{width of } \tfrac{u+v}{2})^2 - (\text{width of } \tfrac{u-v}{2})^2$$

or

$$\text{"correlation"} = \frac{(\text{width of } u+v)^2 - (\text{width of } u-v)^2}{(\text{width of } u+v)^2 + (\text{width of } u-v)^2}$$

as used, for example, by Gnanadesikan and Kettenring (1972). (Linear transformation to avoid large ρ may be important.)

We do not seem to be clear about what specific estimate to recommend. (Comparisons of a variety of resistant/robust estimators by Devlin, Gnanadesikan, and Kettenring are scheduled to appear in the December, 1975 issue of Biometrika.) However, almost any such estimate will be much safer than the classical estimates in the presence of outliers.

The usefulness of taking

$$w = (1-\text{oct})^2$$

where

$$\text{oct} = \max\left\{\left(\frac{u-u'}{cS_u}\right)^2, \left(\frac{v-v'}{cS_v}\right)^2, \left(\frac{z_+-z_+'}{cS_+}\right)^2, \left(\frac{z_--z_-'}{cS_-}\right)^2\right\}$$

and where $z_+ = u+v$, $z_- = u-v$, S_+ is derived from the z_+-z_+', and S_- is derived from the z_--z_-', for some larger values of c, perhaps between 10 and 12, seems not to have been explored.

VECTOR DISPERSION AND CORRELATION AMONG SEVERAL QUANTITIES

DISTINGUISHED CASE

We can assess resistant/robust widths for each coordinate separately, and then divide out by these estimates, coordinate by coordinate, thus obtaining new vectors whose dispersion ought, we would think, be best described by "correlation matrices", by matrices with one's on the diagonal. Thus we can fairly easily, in the distinguished case, reduce the estimation of vector dispersion to the estimation of the various correlations among several quantities.

Devlin, Gnanadesikan and Kettenring have work in progress in this direction which seems likely to offer a variety of approaches, all probably reasonably robust of efficiency. In due course we should be able to identify one or two recommended methods.

THE AFFINE CASE

If we desire an affine invariant estimate of vector disperions, we cannot reduce the problem trivially to the estimation of a correlation matrix. Methods of iterative ellipsoidal truncation and winsorizing have, however, been shown, in unpublished work of Devlin, Gnanadesikan and Kettenring, to be promising.

It may be a little while before this area has been brought into reasonable order.

REFERENCES

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W., Robust Estimators of Location: Survey and Advances, Princeton University Press, 1972.

Andrews, D. F., "A Robust Method for Multiple Linear Regression," Technometrics, 16 (1974), 523-531.

Beaton, A. E. and Tukey, J. W., "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data," Technometrics, 16 (1974), 147-185.

Chen, H., Gnanadesikan, R., and Kettenring, J. R., "Statistical Methods for Grouping Corporations," Sankhyā B, 36 (1974), 1-28.

Day, N. E., "Estimating the Components of a Mixture of Normal Distributions," Biometrika, 56 (1969), 463-474.

Devlin, Susan J., Gnanadesikan, R., and Kettenring, J. R., "Robust Estimation and Outlier Detection with Correlation Coefficients," Biometrika, 62 (1975), 531-547.

Gnanadesikan, R., and Kettenring, J. R., "Robust Estimates, Residuals and Outlier Detection with Multi-Response Data," Biometrics, 28 (1972), 81-124.

Gross, A. M., "A Robust Confidence Interval for Location for Symmetric Long-Tailed Distributions," Proc. Nat. Acad. Sci., 70 (1973), 1995-1997.

Gross, A. M., "Confidence Interval Robustness with Long-Tailed Symmetric Distributions," to appear in J. Amer. Statist. Assoc. for June 1976.

Gross, A. M., "Confidence Intervals for Bisquare Regression Estimates," to be submitted to J. Amer. Statist. Assoc., 1976.

Hampel, F. R., "Robust Estimation: A Condensed Partial Survey," Z. Wahr. verw. Geb, 27 (1973), 87-164.

Hampel, F. R., "The Influence Curve and Its Role in Robust Estimation," J. Amer. Statist. Assoc., 69 (1974), 383-393.

Hill, R. W. and Holland, P. W., "A Monte Carlo Study of Two Robust Alternatives to Least Squares Regression Estimation," NBER Working Paper No. 58 (cited with author's permission), 1974.

Hoaglin, D. C., Holland, P. W., and Welsch, R. E., "Confidence Regions for Robust Regression Estimates," Paper presented to Annual Meeting of American Statistical Association, Atlanta, August, 1975.

Huber, P. J., "Robust Statistics: A Review," Amer. Math. Statist., 43 (1972), 1041-1067.

Huber, P. J., "Robust Regression: Asymptotics, Conjectures and Monte Carlo," Ann. of Statist., 1 (1973), 799-821.

Lax, D. A., "An Interim Report of a Monte Carlo Study of Robust Estimators of Width," Technical Report No. 93, (Series 2), Department of Statistics, Princeton University, 1975.

Tukey, John W., Exploratory Data Analysis, First Edition, to appear Fall 1976.

Velleman, P. F., "Robust Non-linear Data Smoothing," Technical Report No. 89, (Series 2), Department of Statistics, Princeton University (AEC), 1975.

Yuen, K. K. and Dixon, W. J., "The Approximate Behavior and Performance of the Two-Sample Trimmed t," Biometrika, 60 (1973), 369-374.

Yuen, K. K., "The Two-Sample Trimmed t for Unequal Population Variances," Biometrika, 61 (1975), 165-170.

NOSTALIGIC RUMINATIVE CEREBRATIONS

(A Post-Prandial Presentation)

Roger H. Moore
U.S. Nuclear Regulatory Commission
Bethesda, Maryland

My wife and I feel a great pleasure at being invited here tonight. We usually don't practice astrology or numerology--but this date and this place have very special meanings to us. Just 10 months ago tonight, we left the Hill (Los Alamos) and headed for an almost complete change in our way of life--this time in the far and, to us, almost inscrutable East Coast. And it was only 268 months ago tonight that we settled into our first very own place after enjoying two days of mid-1950's old Lodge hospitality. Granted, "our first very own place" was a concrete no-bedroom Iris Street efficiency apartment. But it seemed like a palace after those years of hallowed graduate student poverty level living and a long midsummer desert trek from Oregon in a '47 Kaiser that suffered from an incredible assortment of circulatory and respiratory ailments.

The last time I addressed a group after dinner occurred a little over a year ago at a meeting of the Albuquerque section of the American Society of Quality Control. I was paired with Bill Mondt who then was looking to his first year as head football coach at the University of New Mexico. Surely, what has happened to the Lobos of UNM since then is not a bellwether for the careers of Dr. Tukey and the other invitees at this symposium.

But a certain freedom is granted an afterdinner speaker that is not afforded those making scientific pronouncements. And that's why it was so easy to accept the invitation from Ray Waller when he offered it last spring. Besides, the meeting was then months away. Even in July, when pressed for a title, I willingly accepted a helpmate's help in arriving at the subtly subliminal appellation attached hereto. (Incidentally, Eileen had suggested Noetic as the first word--but I rejected that on the basis that the first week of November is too early for yuletide references.)

By late September, however, a poorly controlled panic set in. What could I bring to this distinguished assembly that had not been said before? Was some startlingly new insight just beyond my grasp? How was I to cope with a title that once seemed so felicitous and now was looming like an albatross? Maybe, if I waited just one more day....

Procrastination has its rewards. In a studied effort to ignore the waiting typewriter and the notes I had collected, I settled down beside them to examine the October 1975 issue of Smithsonian. It carried an excerpt from All the Strange Hours, the autobiography of Loren Eiseley, the famed "literary naturalist and scientist." He tells of the "poignant work of tampering with prehistory," describing how "... some fine day, the kaleidoscope through which we peer at life shifts ... and everything is reordered" and illustates it with a tale about being lost in a Carlsbad sinkhole. He was delivered from his predicament by retracing his steps over fragments of broken stalagmites, crystals never before touched or seen by mankind. He writes:

> "By the time I stood at the cave entrance I was looking at life, at my companions, at the traffic below on the road, as though I had just arisen, a frozen man, from a torrent of melting ice. I wiped a muddy hand across my brow. The hand was 10,000 years away. So were my eyes, so would they always be, and ... I could not find a way to speak.
>
> "The modern world was small, I thought, tiny, constricted beyond belief. A lost century, a toy ....
>
> "I have never again seen men so minutely clear ... moving across an infinite ice field into whose glare we finally vanish. It was like a glimpse through the slitted bone with which Eskimos protect their eyes from snow blindness. I have never had occasion in the years since to think upon us differently. Not once."

Eiseley's eloquence makes things much easier for me. I have a number of ideas to mention tonight--but I am shorn of any pretense that they need be earthshaking. Indeed, that realization makes my joy unconfined. I can now make some chips and see where they fall.

THE AEC AND THE ASG AND THE NRC AND ME

Just so there's no mistake about it at this point in time, let me make an initial point: My entire statistical career has been associated with the Atomic Energy Commission's programs and those of its successors, the Nuclear Regulatory Commission and the Energy Research and Development Administration. Successes and failures, joy and despair, hopes and fears, dreams and realities--for me all of these are tied up with these three agencies. So perhaps you will suffer me whatever "Dutch uncleisms" arise in the sequel.

As I gradually came to an appreciation of the depth and breadth and importance of AEC programs in the 1950s, I also came to a feeling of concern that the discipline we call statistics was under-represented at AEC Headquarters. And at the same time I was buoyed by the discovery that pockets of statistical skill existed among the AEC's laboratories, even if that skill was ill-distributed. But, I wondered, how could a spectrum of statistical questions gain a proper appreciation "in Washington" unless there were a centralized statistical capability in the AEC, clearly identified, and active in dealing with those questions?

I never did find an answer to that question. But I was most surely pleased when, in the late summer of 1974, the Commission approved the formation of an Applied Statistics Group (ASG) within the then-Directorate of Regulation with a clearly-stated "mission" to:

> "1. Provide first-class mathematical statistics, consulting, and policy service to all of REG, including the Director of Regulation and all three Directorates [then called Licensing, Regulatory Standards, and Regulatory Operations]. This includes both solving problems and coming up with new ideas on the application of statistical techniques to regulatory problems.
> "2. Provide a small centralized group for risk evaluation...."

Of course, the Energy Reorganization Act of 1974 was well in the works by then--but the ultimate mission of the ASG could not be diminished by the final details of the Act. As things finally worked out, I boarded the disappearing AEC ship on January 9, 1975, with the understanding that the NRC would effect a rescue on January 19, the effective date of existence for both the NRC and the ERDA. My 10-day AEC tenure set a new record for personal insecurity--with the enlightening exception of a highly instructive ditch-digging engagement in the summer of '45.

ABOUT THE NRC

What about the NRC? What about its philosophy, authorities, responsibilities, activities, functions, and organization. Most of what follows appears in testimony presented by Chairman William A. Anders before the Subcommittee on Energy and Environment, House Committee on Interior and Insular Affairs, April 28, 1975. The full testimony is worth reading, especially as an example of a new government agency's statement of purpose.

Chairman Anders states: "The most basic point is this: The business of Nuclear Regulatory Commission is regulation. We must maintain a position above the fray, and not allow ourselves to be either apologists for or antagonists of nuclear power. Development or promotion of nuclear energy is the function and responsibility of others. Our job is protecting the health and safety of the public."

To guide the Commission, four cardinal principles were adopted: Independence, Openness, Efficiency, and Effectiveness.

The Commission's authorities derive from three statutes: The Atomic Energy Act of 1954 (as amended), the Environmental Policy Act of 1969, and the Energy Reorganization Act of 1974.

The responsibilities of the Commission are protection of the public health and safety, protection of environmental quality, safeguarding nuclear materials and facilities, and ensuring conformity with the nation's antitrust laws.

Activities cover most of the steps of the nuclear fuel cycle, from milling of source materials to either final storage and disposition (wastes) or reprocessing and further use (nuclear reactor fuel and radioisotope byproducts). The NRC does not regulate uranium mining or government enrichment plants.

The functions, described under Title 10 of the Code of Federal Regulations, are supported by broad programs in standards-setting, technical reviews and studies, licensing actions, inspection and enforcement, evaluation of operating experience, and research.

The NRC organization, shown in Figure 1, has been set up to accomplish these functions. Briefly, the five-man Commission is appointed by the President. The Executive Director for Operations (EDO) is the coordinating and directing agent for the Commission. Five major operating offices provide the day-to-day action: Nuclear Reactor Regulation (NRR), Nuclear Material Safety and Safeguards (NMSS), Nuclear Regulatory Research (RES), Standards Development (SD), and Inspection and Enforcement (IE). (Initials included to assist in further encounters with NRC units.)

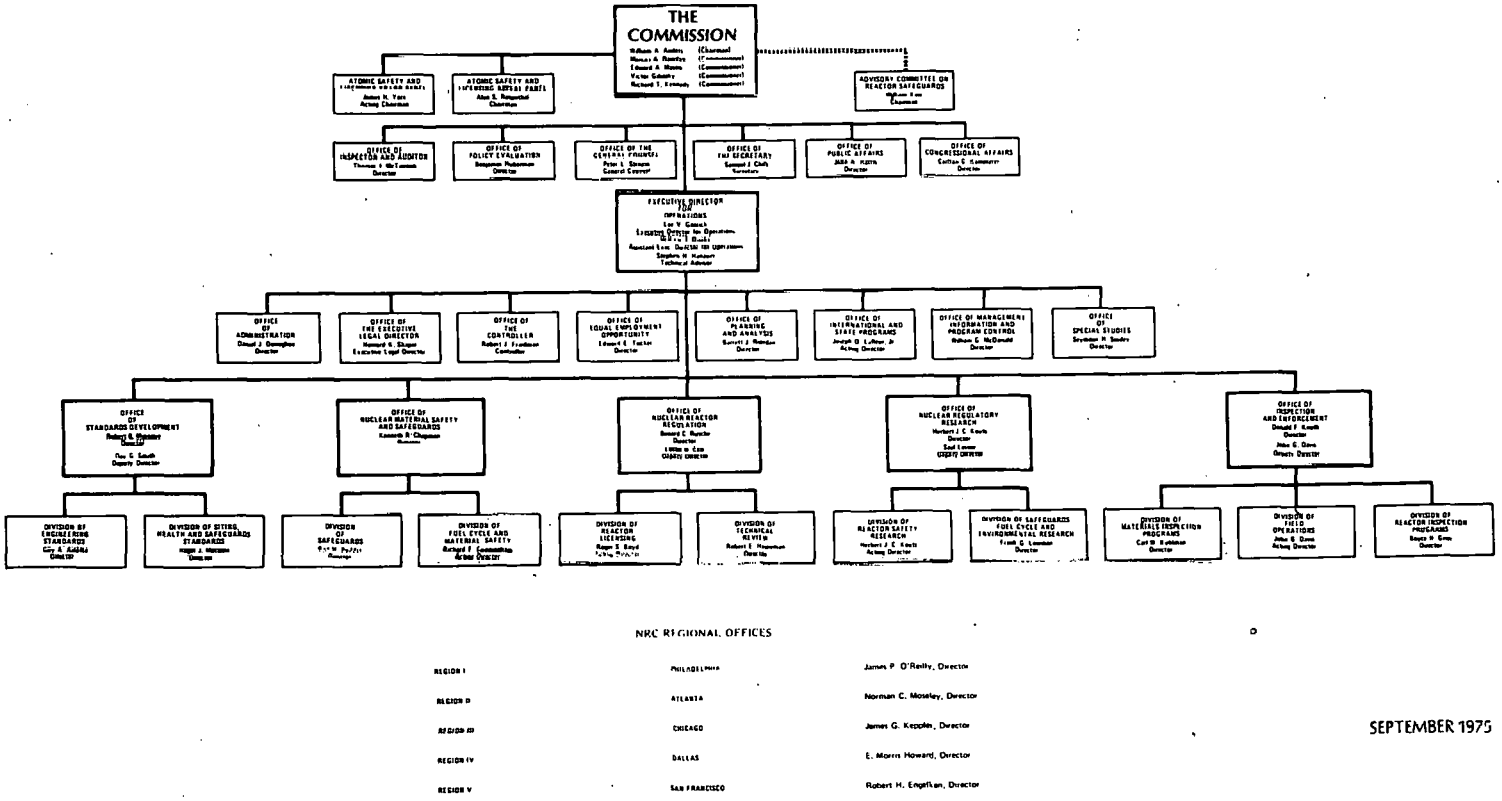


FIGURE 1. Nuclear Regulatory Commission

I won't pretend an intimate acquaintanceship with the myriad boards and associated information that could be cited to flesh out this brief survey. My only intent is to provide some insight into what is happening in the NRC and how things are organized. But I can't resist a couple of statistics: The FY 1976 budget called for 2339 personnel and funds amounting to $220 million (including contractual support).

ABOUT THE ASG

The Applied Statistics Group doesn't appear on the organization chart. It's not organized as an "Office" and it has not been called a "Branch." But "Group" conveys both a sense of upbeat modernity and a feeling of unity of purpose.

Officially, the ASG is responsible to Dr. Stephen H. Hanauer who is the Executive Director's Technical Advisor. Steve has been especially helpful in establishing the ASG's focus for statistical activity in the NRC--sometimes by providing introductions to various segments of the organization, sometimes by forcing us to sharpen our statistical wits in order to parry his queries.

We had occasion recently to recount the Group's endeavors. We were astonished at the number of individual projects to which we had made contributions with some 18 people-months of effort in about eight calendar months. They include studies of fuel element behavior, nuclear materials safeguards information systems, connections among inspection and enforcement and reactor experience, reviews of Regulatory Guides and regulations, analysis of data pertaining to aircraft-carried materials, surveillance of nuclear reactor structures, preparation and review of ANSI-sponsored standards. Even now we are working on a sampling scheme for the Office of the Controller to apply to the auditing of travel vouchers.

The second part of the ASG mission, the risk evaluation effort, was obviated when several members of the Reactor Safety Study (WASH-1400) were brought together in the Office of Nuclear Regulatory Research to form the Probabilistic Analysis Branch, now under the direction of Ian Wall. The skills and training represented in that Branch also are directed at providing reliability theory and practice for NRC programs.

Without people, of course, none of this would be possible. ASG people deserve spotlighting. "Cookie" Ong has been around the licensing and regulatory wars longer than any of us; he joined the ASG in April and his safeguards knowledge is always on call by new arrivals and old hands at the NRC. Bob Easterling is well into a two-year hitch as the ASG's Statistical Advisor; he arrived in June and his contractor-inspired views provide a fine counterpoint with which to probe the verities of nuclear statistics. David Rubinstein is our newest permanent staff member; he joined us in October and his industrial and governmental experience will provide the ASG with its share of the added dimension now so popular with so many institutions. And we must not slight Nancy Barr's secretarial contributions nor her insights into governmental paper management.

Bob Easterling expounds a thesis: "The essense of good regulation is statistics." He and I, most assuredly, have encountered arguments on that--but its usefulness lies in reminding everybody that data, and its analysis and interpretation, ought to underpin the strategies and tactics employed to carry out a regulatory philosophy--irrespective of the thing being regulated. And we recognize that "ought to" provides us with a goal whose achievement will forever seem to be just beyond our grasp.

It is important to note that the dismay I expressed earlier about the under-representation of statistics in the AEC was partially assauged this past summer by an almost singular event. Six senior AEC staff members were honored with Distinguished Service Awards. During the reading of one of the citations, we heard the words "statistics" and "statistical" several times. It's hard to reject such an indicator, no matter what level of significance is chosen.

The conjunction of the authorization of the Applied Statistics Group and the formation of the Nuclear Regulatory Commission and the consequent organization choices provide us with a very special opportunity. Modesty alone cannot account for the feeling of challenge and excitement that goes with being a part of this venture....

A STATISTICAL POTPOURRI

Sometime between the ages of 25 and 79 years, most of us begin each day with a few more unanswered questions than we had the morning before. And it's not completely explained by the action of alcohol on little grey cells. In preparing this talk I spent some time with recent issues of The American Statistician. Reading them sequentially, in a single sitting, is an exercise I highly recommend. Articles therein, along with other signals from many unidentifiable sources, lead me to offer the following for your consideration. (References provided on request.)

Is a statistician a scientist or a shoe clerk?

What is interaction?

Should there be a certification or licensing of statisticians?

How do we cope with a well-established statistician's elementary text that speaks of "2-sigma" confidence intervals?

Is statistics identically equal to data analysis?

Can uncertainty be expressed as a single number?

When can we expect the first "Chemistry for Statisticians" book prepared by a statistician?

Can a systematic error be random? Indeed, can it even have a random component?

Are probability, statistics, and reliability distinguishable disciplines?

If the formulas for the end points of a confidence interval are symbols for random variables, what do we call the realizations computed from the sample data?

Should statistical activists rally around "The Ubiquity of Statistics" thesis enunciated by William Kruskal?

Should we take cognizance of consonance?

If "statistical science training" is good for government employees, is it also good for other people?

If mathematics is the "handmaiden of science," then is statistics not both the taskmaster and stableboy of human endeavor?

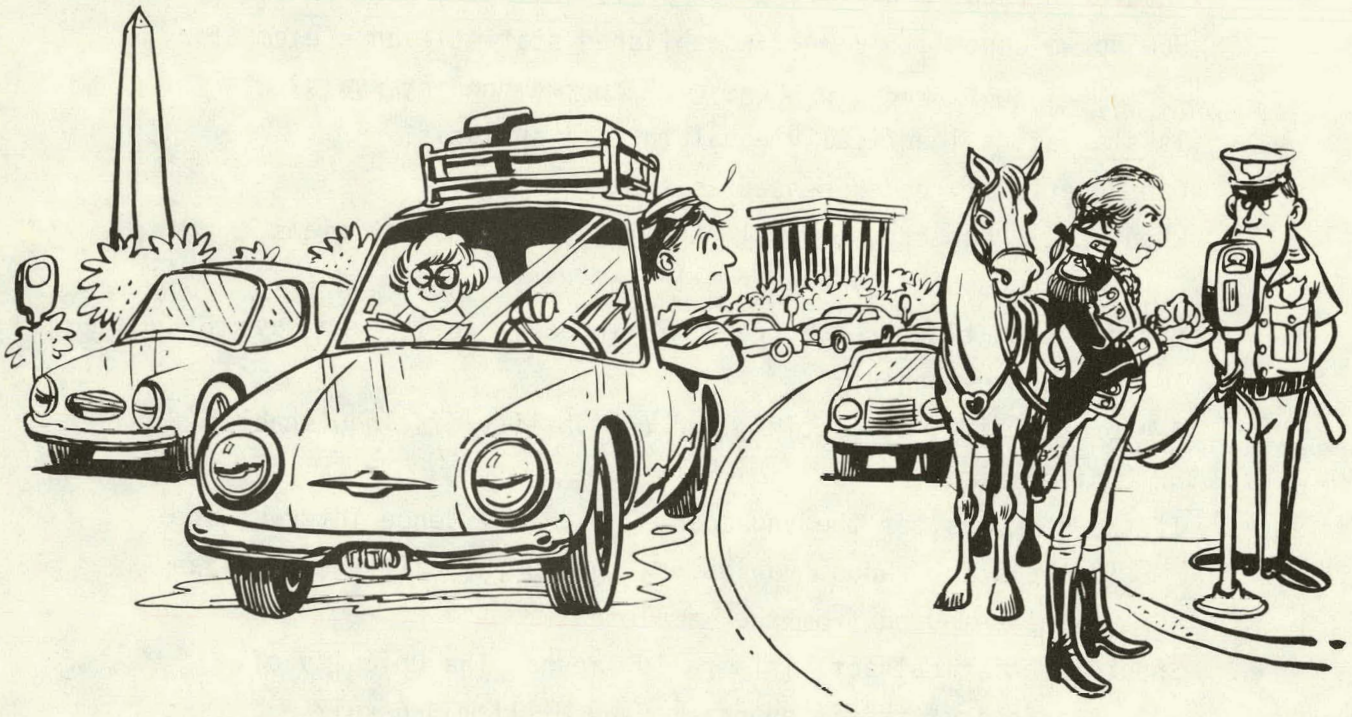
If I possessed comforting answers to these questions, I wouldn't have asked them. Filling the chinks in any system is arduous. But it can

result in stylish and important energy-saving actions. I submit that statistics and statisticians have some work to do.

PARTING SHOT

Our country is in the midst of its Bicentennial Celebration. Washington, D.C. is a splendiferous place to visit any time--now it is doubly so. For those planning to drive, I have a special, but clearly unofficial, offer: Visitors will receive no parking tickets--all the possible parking places will be taken long before anybody arrives from out of town!

Thank you again for inviting us.



RESEARCH ON DATA ANALYSIS TECHNIQUES AT STANFORD LINEAR ACCELERATOR CENTER

(An After Dinner Presentation)

Jerome H. Friedman
Stanford Linear Accelerator Center
Stanford, California

INTRODUCTION

To understand the approaches we have taken to data analysis techniques at Stanford Linear Accelerator Center (SLAC), it is necessary to understand the types of data that we typically encounter in high energy particle physics experiments. These data originate with the particle detectors associated with scattering experiments. These detectors produce counts whenever a particle passes through them. The cardinality of these counted data tend to be larger than that usually encountered by statisticians. Sample sizes from 5,000 to 50,000 are not uncommon. The data are nearly always multivariate with typical dimensionalities ranging from 4 to 20. The data distributions tend to be far from normal. These distributions are highly structured in the multivariate data space, and location is not always their most important attribute. Quite often the scale of a distribution is a much more important characteristic than location.

These characteristics place severe restrictions on the nature of the procedures that we employ. They also allow for some exciting possibilities. The non-normality of the distributions requires that we employ nonparametric procedures. The fact that location is not the most important attribute of the distributions rules out most of the nonparametric techniques that are simply robust extensions of normal procedures. Totally general nonparametric techniques are required. The multi-attribute nature of the data requires that these procedures be multivariate. The large sample sizes require that these techniques be computationally efficient. Procedures in which the computation time grows as the square of the sample size, or faster, are unacceptable. On the other hand, the large sample sizes allow for the possibility that we at least have a chance of being successful with generalized nonparametric multivariate techniques. The so-called

"curse of dimensionality" requires that the sample size, N, grow with dimension, p, as N^p in order to maintain roughly constant precision. Thus, very large sample sizes are required for even moderate dimensionalities.

With parametric techniques, the data distribution is assumed to be described by a small number of parameters that are to be estimated. This effectively reduces the dimensionality to that of the parameter space, considerably alleviating the problem. In our applications, however, models describing the densities are generally not known, ruling out this approach. In fact, an important problem is validating models that are proposed by theoretical physicists, given a multivariate data sample. Although described in a parametric framework, goodness-of-fit testing is really a nonparametric problem.

Our first efforts were directed toward searching the literature for procedures that met the requirements described above. We found two independent and largely noninteracting bodies of literature addressing themselves to these types of problems. The first was that of statisticians and their adherents (biometrics, econometrics, social sciences, education, etc.). The second was that of electrical engineers and their associates (pattern recognition, artificial intelligence, computational geometry, etc.)

We found very little that we could apply in the general statistics literature. Almost all multivariate procedures were based on normality. Those nonparametric procedures that did exist for multivariate applications were usually robust generalizations of normal procedures (for example, replacing datum locations by corresponding ranks). The most promising techniques that we did find were those based on tolerance regions. On the other hand, the statisticians tended to take considerable care in deriving the expected performance of their procedures, at least in the asymptotic limit. Good asymptotic performance is a necessary but usually not a sufficient guarantee of a technique's usefulness. This is because the "curse of dimensionality" can require that the sample size become truly astronomic before asymptotic results are valid.

We found a great body of work in the engineering literature that approaches problems similar to ours, some of which we have found to be quite useful. The principal drawbacks to much of this work are lack of demonstrated performance and excessive computational complexity. These procedures tended to be completely heuristic. Their performance was usually demonstrated by application to a single specific problem (usually two-dimensional) along with speculation as to the general performance (both statistically as well as computationally). This speculation usually turned out to be incorrect, as we found by trying them on more general problems and at higher dimensionality. However, we did at least find a wealth of innovative ideas in this literature.

We learned two lessons from this exercise. First, the state of the art would probably be much more advanced today if there had been more communication between these two bodies of literature. Second, if we wanted solutions to our problems we would, for the most part, have to find them ourselves. As our efforts progressed, however, we began to understand the reasons for the lack of advancement in this field. We soon came to realize that the problems we had set out to solve were quite difficult. Our approach has been two-fold. First, to try to find ways of reducing the computational effort of techniques that otherwise look promising, and second, try to develop new procedures that fit our needs. In spite of the difficulty of the problems, we have made some progress.

REDUCING COMPUTATION COMPLEXITY

The class of existing techniques that appeared most promising to us were those based on interpoint distances. At the heart of the analysis of counted data is local density estimation. Density can be estimated by fixing a small region in the multidimensional data space and counting the number of data points that lie within the region. The estimate is then the number of points divided by the volume (Rosenblatt, 1956, or Parzen estimator, 1962). An alternative is to fix the number of data points and measure the volume that exactly contains that number (Loftsgaarden and Quesenberry, 1965). A great many general nonparametric techniques are based either directly or indirectly on this type of density

estimation. In fact, many procedures require such an estimate for each data point. Computationally, these procedures require searching the multidimensional data space for data points that are close to a given prespecified point. From the computer science point of view, this is a problem in computational geometry or searching on secondary keys. At the time we began considering these problems, the state of the art was simply to calculate the distance to all points from the prespecified point, identifying those that were smallest. This required computation proportional to the sample size, N. If this was to be done for all data points, the computation increased with increasing N, as N^2. This limited the application of these procedures to small data samples. One of our principal accomplishments has been to develop an algorithm that performed this searching with computation proportional to NlogN (Friedman, et al., 1975). This reduced, by many orders of magnitude, the computation required to obtain density estimates using interpoint distances and greatly extended the applicability of these procedures to large data sets.

A somewhat related problem is that of shortest connecting networks or minimal spanning trees. A minimal spanning tree for a set of points in a multidimensional space is a network of links or branches connecting them that have minimum total distance while providing a route between every pair of points. Methods based on minimal spanning trees have been proposed for cluster analysis and data summarization [Zahn (1971), Arkadev and Braverman (1967), Johnson (1967), and Gower and Ross (1969)]. We have found that these graph theoretical methods are among the most powerful for detecting and describing clustering in multivariate data. The problem was that the fastest known algorithms for constructing minimal spanning trees required computation proportional to N^2 [Kruskal (1956), Prim (1957), and Dijkstra (1959)]. This computation limited these techniques to small data samples. Using the fast searching algorithms that we developed for density estimation, we were able to develop an algorithm for constructing minimal spanning trees with computation proportional to NlogN (Bentley and Friedman, 1975). This greatly expands the applicability of these clustering techniques to much larger data bases.

NEW PROCEDURES IN MULTIVARIATE DATA ANALYSIS

Our research on new procedures for multivariate data analysis has centered on both exploratory and confirmatory techniques. The exploratory techniques were in large part developed in collaboration with Professor John W. Tukey of Princeton University and Bell Laboratories at Murray Hill. These methods center on mapping techniques. That is, the p-dimensional data is mapped onto one, two or perhaps three-dimensional manifolds in a manner that preserves, as closely as possible, some aspect of the data structure in the full dimensionality of the data space. The resulting data distributions are viewed by a researcher in this lower dimensionality where the human gift for recognizing data relationships as patterns can be employed.

Similar techniques for this purpose have been separately proposed in both the statistics and engineering literature. Statisticians refer to these as "multidimensional scaling" while the engineers refer to them as "nonlinear mapping". When they can be applied, these techniques can be quite powerful. They suffer from two important limitations. The first is astronomical computational complexity and the second is difficulty in interpreting the resulting maps.

We have termed our approach "projection pursuit". As the name implies, we employ projections or linear mapping techniques. The advantage of linear maps is ease of interpretation and computational simplicity. The disadvantage is that when generally applied, low dimensional projections usually preserve very little of the multivariate structure of the data. The purpose of projection pursuit is to use the multivariate data set to find those projections that are most revealing of the multivariate data structure.

We have implemented this concept in two quite different ways. One is on an interactive computer graphics device. We have named this device PRIM-9 (Steppel, 1973, and Fisherkeller, 1974). It provides the researcher with ability to rotate the data to any desired orientation, while continuously viewing a two-dimensional projection of the multivariate data. These rotations are performed in real time and in a continuous manner

under operator control. This gives the researcher the ability to perform manual projection pursuit. That is, by controlling the rotations and viewing the changing projections, he can try to discover those data orientations (or equivalently, projections) that reveal to him interesting structure.

The other implementation of this concept was as an automatic computer algorithm (Friedman and Tukey, 1974). This algorithm automatically searches the data measurement space for revealing one and two-dimensional projections. These projections can then be inspected by the researchers for interpretation.

We have also made some progress in the area of confirmatory data analysis techniques. We developed a general test for the stochastic independence of two multivariate data samples (Friedman, 1974). We have also developed a general nonparametric procedure for testing whether two multivariate point distributions are drawn from the same probability density function (Friedman, et al., 1973). As well as testing if the two sample distributions are different, it also can identify the regions of the multivariate data space where the two distributions most differ, and where they are most similar.

We have developed a whole class of procedures based on the concept of data driven recursive partitioning. This technique goes as follows: Consider a set of multivariate observations, S_0. Based on these observations choose by some procedure a direction, $\hat{d}$, and a scalar split point, s. The sample S_0 is then divided into two subsamples S_1 and S_2 such that

$$\vec{x} \in S_1 \text{ if } \vec{x} \cdot \hat{d} \leq s$$
$$\vec{x} \in S_2 \text{ if } \vec{x} \cdot \hat{d} > s \ .$$

This partitioning procedure is then reapplied to the subsamples S_1 and S_2, obtaining four subsamples; each of these are then partitioned similarly, and so on. The recursive application of this partitioning procedure is continued until all subsets finally meet a terminating condition. These terminal subsets represent mutually exclusive subsamples of the complete data sample, and also represent convex subregions (called buckets) that collectively cover the data measurement space.

A specific multivariate procedure is defined by providing the prescription for determining the partitioning direction, $\hat{d}$, and the split point, s, for partitioning each subsample, and the terminating condition that ends the partitioning. These prescriptions are usually cast in the form of optimizing some figure of merit. The goal of the partitioning, and thus, the figure of merit, is to configure the terminal buckets so that data contained in the same bucket share as closely as possible some common property.

Computationally, this data partitioning can be represented by a binary tree. Binary trees are well known to have nice computational properties. In particular, they can usually be constructed in O(NlogN) time and be searched in O(logN) time. By providing various prescriptions for the partitioning criteria, we have applied this technique to a wide range of problems. A "variable metric" decision rule for nonparametric classification has been developed (Friedman, 1975). This rule is invariant under monotone transformations of the measurement coordinates; its performance is unaffected by the presence of coordinates that contain little or no discriminating information, and it handles data with missing measurements with minimal information loss. We have applied this technique to nonlinear regression problems. The recursive partitioning divides the dependent variable space into separate regions such that, within each, the dependence is as linear as possible. It then performs a linear regression in each such region separately. We have also developed an algorithm, based on data-driven recursive partitioning, for estimating intrinsic dimensionality of multivariate data sets and performing generalized outlier analysis.

To summarize, we have found the concept of data-driven recursive partitioning to be a powerful tool for developing effective multivariate procedures with low computational complexity.

CONCLUSION

As the above discussion indicates, we have made some progress toward the goals outlined in the introduction. We have not yet solved some of our more important problems. For example, a general multivariate goodness-of-fit

test that is both powerful and computable has eluded us so far. Effective and computable nonlinear mapping algorithms are not yet available. But, I think the progress that we have made indicates that these problems may not be intractable, and that we may see their solutions soon.

REFERENCES

Arkadev, A. G., and Braverman, E. M., "Computers and Pattern Recognition", Washington, DC., Thompson Book Co., 1967.

Bentley, J. L., and Friedman, J. H., "Fast Algorithms for Constructing Minimal Spanning Trees in Coordinate Spaces", SLAC-PUB 1665, Stanford Linear Accelerator Center, October 1975.

Dijkstra, E. W., "A Note on Two Problems in Connection with Graphs", Numer. Math 1 No. 5, (1959), 269-271.

Fisherkeller, M. A., Friedman, J. H., and Tukey, J. W., "PRIM-9--An Interactive Multidimensional Data Display and Analysis System", Proceedings of Second Annual AEC Scientific Computer Information Exchange Meeting, New York City, (May 2-3, 1974), 3-33.

Friedman, J. H., "Data Analysis Techniques for High Energy Particle Physics", SLAC-176, Standord Linear Accelerator Center, (1974), 73-76.

Friedman, J. H., "A Variable Metric Decision Rule for Nonparametric Classification", STAN-CS-75-487, Stanford University Computer Science, (April 1975) (to be published in IEEE Trans. Comput.)

Friedman, J. H., Bentley, J. L., and Finkel, R. A., "An Algorithm for Finding Best Matches in Logarithmic Time", STAN-CS-75-482, Stanford University Computer Science, (February 1975) (to be published in ACM Transactions on Mathematical Software).

Friedman, J. H., Steppel, S., and Tukey, J. W., "A Nonparametric Procedure For Comparing Multivariate Point Sets", Stanford Linear Accelerator Center, Computation Research Group Tech. Memo No. 153, 1973.

Friedman, J. H., and Tukey, J. W., "A Projection Pursuit Alogrithm for Exploratory Data Analysis", IEEE Trans. Comput., C-23 (1974), 881-890.

Gower, J. C., and Ross, G. J. S., "Minimal Spanning Trees and Single Linkage Cluster Analysis", Appl. Statistics, 18 (1969), 54-64.

Johnson, S. C., "Hierarchial Clustering Schemes", Psychometrika, 32 (September 1967), 241-254.

Kruskal, J. B., Jr., "On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem", Proc. Amer. Math. Soc., 7 (1956), 48-50.

Loftsgaarden, D. O., and Quesenberry, C. P. "A Nonparametric Density Function", Ann. Math. Stat., 36 (1965), 1049-1051.

Parzen, E., "On the Estimation of a Probability Density Function and the Mode", Ann. Math. Stat., 33 (1962), 1065-1076.

Prim, R. C., "Shortest Connection Networks and Some Generalizations", Bell Syst. Tech. J., 36 (1957), 1389-1401.

Rosenblatt, M., "Remarks on Some Nonparametric Estimates of A Density Function", Ann. Math. Stat., 27 (1956), 832-837.

Steppel, S., editor, "PRIM-9", produced by Stanford Linear Accelerator Center, Stanford, CA., Bin 88 Productions, April 1973.

Zahn, C. T., "Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters", IEEE Trans. Comput., C-20 (January 1971), 68-86.

RESEARCH PAPERS

LOWER BOUNDS FOR THE MULTIVARIATE NORMAL MILLS' RATIO

G. P. Steck
Sandia Laboratories
Albuquerque, New Mexico

Lower bounds are derived for the multivariate Mills' ratio by expressing it as an expectation of a convex function and using Jensen's inequality.

INTRODUCTION

Let $\underline{X} = (X_1, X_2, \ldots, X_n)$ be a vector of standardized normal random variables with $EX_iX_j = \rho_{ij}$. Let the positive definite covariance matrix be $\underline{\Sigma}$ with $\underline{M} = \underline{\Sigma}^{-1}$. The multivariate Mills' ratio, $R(\underline{a},\underline{M})$, is defined to be the multivariate normal probability "beyond" a certain cutoff point divided by the multivariate normal density at that point; that is

$$R(\underline{a},\underline{M}) = (2\pi)^{n/2} \, |\underline{\Sigma}|^{1/2} \exp(\underline{aMa}'/2) \, P(\underline{X} \geq \underline{a}).$$

Savage (1962) generalized the Shenton (1954) formula for the univariate Mills' ratio obtaining the representation

$$R(\underline{a},\underline{M}) = \int_{\underline{u}\geq\underline{0}} \exp(-\underline{aMu}' - \underline{uMu}'/2)d\underline{u}. \tag{1}$$

Asuming $\underline{\Delta} = \underline{aM} > 0$, Savage derived upper and lower bounds for R by using $1-x \leq e^{-x} \leq 1$ on $\exp(-\underline{uMu}'/2)$. His bounds are

$$\frac{1 - \sum_1^n m_{ii}/\Delta_i^2 - \sum\sum_{i<j} m_{ij}/\Delta_i\Delta_j}{\Delta_1\Delta_2 \cdots \Delta_n} \leq R(\underline{a},\underline{M}) \leq \frac{1}{\Delta_1\Delta_2 \cdots \Delta_n}.$$

With the same assumption Ruben (1964) carried the expansion to more terms showing that this led to an enveloping asymptotic expansion which gave alternating upper and lower bounds of increasing complexity.

More recently, Gjačjauskas (1973) showed that if $a_i \to \infty$ for some i and $a_i > 0$, all i, then

$$\frac{1}{z} - \frac{1}{z^{\frac{n+1}{n}}} < R(\underline{a},\underline{M}) < \frac{1}{z} + \frac{1}{z^{\frac{n+1}{n}}} ,$$

where $z = (1/2)^n \prod_1^n \partial(\underline{x}\underline{M}\underline{x}')/\partial x_i = \prod_{i=1}^n \left[\sum_{j=1}^n m_{ij}x_j\right]$,

provided $m_{ij} > 0$ all i, j. For n = 1 this result is weaker than the usual one.

If one is interested only in bounding multivariate normal probabilities, there is a result due to Slepian (1962) that may be useful. Slepian showed that $P(\underline{X} > \underline{a} \mid \underline{\Sigma} = \underline{C}) > P(\underline{X} > \underline{a} \mid \underline{\Sigma} = \underline{D})$ provided $c_{ij} > d_{ij}$ ($c_{ii} = d_{ii} = 1$, all i). So, for example, if all the correlations are positive, one could bound $P(\underline{X} > \underline{a})$ above and below by probabilities based on the equicorrelated cases for $\rho = \max_{i,j} \rho_{i,j}$ and $\rho = \min_{i,j} \rho_{i,j}$, respectively.

In this paper we derive three approximations from equation (1), including bounds, two of which are expressed in terms of the univariate Mills' ratio. These approximations and bounds do not require $\underline{\Delta} > \underline{0}$ and are all based on $Ef(\underline{X}) \simeq f(E\underline{X})$, with $Ef(\underline{X}) > f(E\underline{X})$ if f is convex in each argument.

In the last section we report on the results of a sampling study carried out to evaluate the accuracy of the approximations. That study showed the bounds to be reasonably good and much better than the Savage bounds which rarely existed. The requirement that $\underline{\Delta} > \underline{0}$ appears to be a severe one.

APPROXIMATIONS AND BOUNDS

Make the change of variable $u_i = v_i/\sqrt{m_{ii}}$, i = 1, 2, ..., n in equation (1) and rearrange the integrand so that

$$R(\underline{a},\underline{M}) = \prod_1^n (m_{ii})^{-1/2} \int_0^\infty \cdots \int_0^\infty e^{-\underline{v}\underline{P}\underline{v}'/2} \left(\prod_1^n e^{-z_i v_i - v_i^2/2} \right) dv_i , \qquad (2)$$

where $\underline{Q} = \left(m_{ij}/\sqrt{m_{ii}m_{jj}}\right)$, $\underline{P} = \underline{Q} - \underline{I}$ and $z_i = \Delta_i/\sqrt{m_{ii}}$. This manipulation expresses $R(\underline{a},\underline{M})$ as proportional to $E\exp(-\underline{V}\underline{P}\underline{V}'/2)$ where $\underline{V} = (V_1, V_2, \ldots, V_n)$ is a vector of independent random variables such that V_i has a density proportional to $\exp(-z_iv_i - v_i^2/2)$. An elementary computation shows $E_i \equiv EV_i = -R'(z_i)/R(z_i) = E(z_i) = z_i - 1/R(z_i)$, where $R(x) = \int_0^\infty \exp(-xt-t^2/2)dt \equiv R(x,1)$ is the univariate Mills' ratio. Since $\underline{P}$ contains only crossproduct terms, $\exp(-\underline{V}\underline{P}\underline{V}'/2)$ is convex in each argument and the resulting approximation is a bound. Thus we have

$$R(\underline{a},\underline{M}) > \prod_1^n \left[R(z_i)/\sqrt{m_{ii}}\right] \exp(-\underline{E}\underline{P}\underline{E}'/2).$$

In the evaluation this bound is called BOUND.

Another approximation, which is a lower bound for n = 2 and which appears to be a lower bound for n > 2 as well, is obtained from equation (2) by integrating out one variable at a time. The result is

$$R(\underline{a},\underline{M}) \simeq \prod_1^n \left[R(w_i)/\sqrt{m_{ii}} \right],$$

where

$$w_n = z_n$$

$$w_k = z_k + \sum_{i=k+1}^{n} q_{ki}E(w_i), \quad k = n-1, n-2, \ldots, 1$$

and $q_{ij} = m_{ij}/\sqrt{m_{ii}m_{jj}}$, $E(w) = -R'(w)/R(w)$ as before. In the evaluation this approximation is called APPRO. For n = 2, APPRO is a lower bound for $R(\underline{a},\underline{M})$ and

$$APPRO = (1-\rho^2)\max\left\{R(z_1)R[z_2-\rho E(z_1)],\ R(z_2)R[z_1-\rho E(z_2)]\right\}.$$

Since APPRO depends on the order in which the variables are integrated out there are n! different possible values. In the evaluation we considered three possibilities: (1) integration of the variables in their natural order, (2) integration of the variables in an order corresponding to the value of z with the variable with the largest value of z being integrated out first, (3) same as (2) except that it is the variable with the smallest z which is integrated out first. In the cases studied each of these values

of APPRO was a lower bound, so it is possible that the maximum of the n! possibilities is also a lower bound. It is also possible that an algorithm which minimizes the maximum w or maximizes the minimum w would give w's for which APPRO approximated the maximum of the n! possibilities reasonably well.

Another bound can be derived from equation (2) by moving the $\exp(-\underline{z}\underline{v}')$ factor over with $\exp(-\underline{v}\underline{P}\underline{v}'/2)$. Then the expectation with respect to independent normal random variables is restricted to $(0,\infty)$. (Another possibility, of course, is to move the $\exp(-\underline{v}\underline{v}'/2)$ factor and take the expectation with respect to independent exponential random variables, but this would require $\underline{\Delta} > \underline{0}$.)

The function $\exp(-\underline{v}\underline{P}\underline{v}'/2 - \underline{z}\underline{v}')$ is convex in each argument since $\underline{P}$ has zero entries on the diagonal and we have

$$R(\underline{a},\underline{M}) \geq \left(\frac{\pi}{2}\right)^{n/2} \prod_{1}^{n} (m_{ii})^{-1/2} \exp\left[-\sqrt{\frac{2}{\pi}}\sum z_i - \frac{2}{\pi}\sum_{i<j}\sum q_{ij}\right].$$

where, as before, $z_i = \Delta_i/\sqrt{m_{ii}}$ and $q_{ij} = m_{ij}/\sqrt{m_{ii}m_{jj}}$. In the evaluation we have called this bound BDBDB.

EVALUATION

The multivariate normal integral can be expressed as a single integral of products of univariate normal integrals when there is a set of constants $b_1, b_2, \ldots, b_n$ such that $\rho_{ij} = b_i b_j$. In that case $X_i = b_i Y + \sqrt{1 - b_i^2}\ Z_i$ where $Y, Z_1, Z_2, \ldots, Z_n$ are i.i.d. N(0,1) and the X's are conditionally independent given $Y = y$. Letting G denote the standard normal distribution function, it follows that

$$P(\underline{X} > \underline{a}) = \int_{-\infty}^{\infty} \prod_{1}^{n} G\left(\frac{-a_i + b_i Y}{\sqrt{1 - b_i^2}}\right) G'(y)dy .$$

Hence, in this particular case, $P(\underline{X} > \underline{a})$ can be computed by standard quadrature methods and for purposes of evaluating the approximations and bounds, we considered only this special case. The exact Mills' ratio and its approximations were computed for n = 2(2)20 and many combinations of {a} and {b} where the b's were chosen independently from a uniform distribution

on (-1,1) and the a's chosen independently so that a range of values of $P(X > a)$ were obtained. The results are summarized in Table 1.

TABLE 1. Values of $P(X > a)$, Mills' Ratio and Approximations BOUND, APPRO, BDBDB and Savage for n = 2, 4, 6, 10, 20

n	Probability	Mills' Ratio	BOUND Approx.	BOUND %	APPRO Approx.	APPRO %	BDBDB Approx.	BDBDB %	SAVAGE Lower Bd.	SAVAGE Upper Bd.
2	.114391	.8910	.8793	1.3	.8832	0.9	.8716	2.2	-43.5	5.07
	.029434	.5327	.5060	5.0	.5128	3.7	.4922	7.6	-13.3	1.86
	.009851	.4368	.4368	.0	.4368	.0	.2672	38.8	-20.8	1.80
	.001028	.2529	.2460	2.7	.2480	1.9	.2035	19.5	.033	.414
	.000183	1.0942	1.0924	.2	1.0934	.1	.2254	79.4	*	*
4	.104284	2.4349	1.6633	31.7	1.8696	23.2	1.6632	31.7	*	*
	.031878	1.2643	.9212	27.1	1.0194	19.4	.8113	35.8	*	*
	.010031	.7684	.7572	1.5	.7581	1.3	.6238	18.8	*	*
	.001031	.1943	.1921	1.1	.1935	.4	.0939	51.7	-33.6	2.41
	.000183	.0448	.0430	4.0	.0440	1.8	.0171	61.8	-16680.	5.98
6	.014542	3.0269	2.8439	6.0	2.9332	3.1	2.8534	5.7	*	*
	.005129	1.5196	1.2923	15.0	1.3587	10.6	1.0651	29.9	*	*
	.000981	.5455	.5066	7.1	.5290	3.0	.3529	35.3	*	*
	.000537	.1777	.1690	4.9	.1725	2.9	.1133	36.2	*	*
	.(4)928	.0670	.0609	9.1	.0642	4.2	.0254	62.1	*	*
10	.010030	83.86	25.41	69.7	38.08	54.6	17.03	79.7	*	*
	.003145	31.77	29.41	7.4	29.56	7.0	25.49	19.8	*	*
	.001042	3.983	3.342	16.1	3.715	6.7	2.369	40.5	*	*
	.000108	.3352	.2610	22.1	.2884	14.0	.1210	63.9	*	*
	.(4)273	.2258	.2024	10.4	.2160	4.3	.1059	53.1	*	*
20	.001334	149600.	42280.	71.7	58140.	61.1	5471.	96.3	*	*
	.000599	45660.	13580.	70.3	18620.	59.2	2230.	95.1	*	*
	.000304	10390.	3281.	68.4	4480.	59.6	674.4	93.5	*	*
	.(4)496	6089.	4764.	21.8	5378.	11.7	1628.	73.3	*	*
	.(4)210	405.3	143.3	64.6	198.4	51.0	33.9	91.6	*	*

*Savage bounds not appropriate because $\Delta \leq 0$.

It is apparent that APPRO is, in general, a good approximation and is the best of the three. BOUND is almost as good as APPRO and BDBDB is, in general, a poor approximation. Furthermore, the bounds given by Savage are poor the few times they are appropriate, i.e., $\Delta > 0$.

REFERENCES

Gjačjauskas, E., "Estimate of the Multidimensional Normal Probability Law for the Vanishing Hyperangle," Litovsk. Mat. Sb. 13 (1973), 83-90. (In Russian with English summary.)

Ruben, H., "An Asymptotic Expansion for the Multivariate Normal Distribution and Mills' Ratio," J. Res. Nat. Bur. Standards, Sect. B, 68 (1964), 3-11.

Savage, I. R., "Mills' Ratio for Multivariate Normal Distributions," J. Res. Nat. Bur. Standards, Sect. B, 66 (1962), 93-96.

Shenton, L. R., "Inequalities for the Normal Integral Including a New Continued Fraction," Biometrika, 41 (1954), 177-189.

Slepian, D., "The One-Sided Barrier Problem for Gaussian Noise," Bell System Tech. Jour., 41 (1962), 463-501.

RANDOM MATRICES AND RANDOM DIFFERENCE EQUATIONS

V. R. R. Uppuluri
Union Carbide Corporation Nuclear Division
Oak Ridge, Tennessee

In a discrete time model and a one-compartment system, statistical distribution of the amount of a substance after m independent trails can be obtained. Superimposition of randomness helps provide confidence bands around the expectation curves. These ideas are generalized to multicompartment models with random matrices. In a mammillary system with special probability structures, closed forms for some distributional characteristics are obtained. Model feasibility in terms of parameters introduced is studied in connection with data on strontium-90 retention.

INTRODUCTION

In this paper we discuss mathematical models leading to products of random matrices and random difference equations. We introduce a one-compartment model with random behavior and show how the average concentration in the discrete time model converges to the exponential function. These ideas are then generalized to two-compartment models and mammillary systems, where products of random matrices appear in a natural way. It is briefly mentioned how the products of random matrices appear in applications in demography and control theory. Random sequences motivated from the following problems are studied:

1) constant pulsing and random decay models,
2) random pulsing and constant decay models, and
3) random pulsing and random decay models.

We discuss some directions for future research in the following topics: applications to genetic evolution, applications to record values, applications to Ising models, random operators and random fixed point theorems.

ONE-COMPARTMENT MODEL

In blood-bone systems it is of interest to understand how radioactivity gets trapped in the bone structure. As one possible approach, stochastic compartmental analysis was developed. In this section we develop a discrete time model and a one-compartment system with a valve that opens and closes at random. In later sections we extend these results to two-compartment models and mammillary systems.

In first order kinetic models, the concentration of a substance, $C(t)$, as a function of time is described by

$$dC(t)/dt = -\lambda\, C(t) \text{ or } C(t) = C(0)\, e^{-\lambda t}. \tag{1}$$

For more details, see Jacquez (1972). We set up a stochastic phenomenon governing the dynamic behavior of the system and show that the average concentration satisfies the above exponential model.

Let the time interval $(0,t]$ be divided into m equal parts so that $t = m\cdot\Delta t$. For convenience let $C(0) = 1$, $a = e^{-\lambda\cdot\Delta t} = e^{-\lambda t/m}$. Let C_m denote the concentration at $m\cdot\Delta t$. Let X_m denote a sequence of independent identically distributed Bernoulli random variables with probabilities,

$$P[X_m = 1] = p \text{ and } P[X_m = 0] = 1 - p = q.$$

If $X_m = 1$, let $C_m = a\, C_{m-1}$ and if $X_m = 0$ let $C_m = C_{m-1}$; or

$$C_m = a\, C_{m-1}\, X_m + C_{m-1}(1-X_m). \tag{2}$$

Clearly, C_m takes one of the values of the set $(1, a, \ldots, a^m)$ and

$$P[C_m = a^k] = \binom{m}{k} p^k q^{m-k},\ k = 0, 1, 2, \ldots, m. \tag{3}$$

The expected value of C_m is equal to

$$E[C_m] = (q + a\, p)^m = (q + p\, e^{-\lambda t/m})^m. \tag{4}$$

Taking limit of $E[C_m]$ as $m \to \infty$, $\Delta t \to 0$ such that $m\,\Delta t \to t$, we obtain

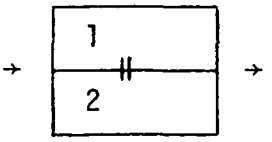
$$\lim E[C_m] = e^{-p\lambda t}. \tag{5}$$

When $p = 1$, we have the result corresponding to the deterministic process.

The superimposition of randomness on the valve helps us to study the behavior of the concentration at time $m\Delta t$ in a more detailed manner. For instance, if we know p, we can obtain the confidence limits on C_m.

TWO-COMPARTMENT MODEL

Suppose we have two compartments of the same volume with a valve that can be opened or closed at will. Suppose there is provision for an inlet and an outlet through compartment 1 which will enable the concentration (in compartment 1 only if the valve is open) to reduce by a factor of $e^{-\lambda\Delta t} = a$, in a time interval Δt.



Let $C_m = \begin{bmatrix} C_{1,m} \\ C_{2,m} \end{bmatrix}$ denote the vector of concentrations in compartments 1 and 2 at time $m\Delta t$. Let X_m denote a sequence of independent identically distributed Bernoulli random variables with $P[X_m = 1] = p$ and $P[X_m = 0] = 1 - p = q$. If $X_m = 1$, let

$$C_m = \begin{bmatrix} (C_{1,m-1} + C_{2,m-1})\frac{a}{2} \\ (C_{1,m-1} + C_{2,m-1})\frac{a}{2} \end{bmatrix}, \text{ and}$$

if $X_m = 0$, let

$$C_m = \begin{bmatrix} C_{1,m-1}\ a \\ C_{2,m-1} \end{bmatrix} \quad \text{or}$$

$$C_m = \begin{bmatrix} C_{1,m} \\ C_{2,m} \end{bmatrix} = \begin{bmatrix} (C_{1,m-1} + C_{2,m-1})\frac{a}{2}X_m + C_{1,m-1}\ a(1-X_m) \\ (C_{1,m-1} + C_{2,m-1})\frac{a}{2}X_m + C_{2,m-1}\ (1-X_m) \end{bmatrix}$$

$$= \begin{bmatrix} a - \frac{a}{2}X_m & \frac{a}{2}X_m \\ \frac{a}{2}X_m & 1 - (1 - \frac{a}{2})X_m \end{bmatrix} \begin{bmatrix} C_{1,m-1} \\ C_{2,m-1} \end{bmatrix}$$

$$= \Gamma_m C_{m-1} = \Gamma_m \Gamma_{m-1} \cdots \Gamma_1 C_0 , \qquad (6)$$

where

$$\Gamma_j = \begin{bmatrix} \frac{a}{2} & \frac{a}{2} \\ \frac{a}{2} & \frac{a}{2} \end{bmatrix} = A_1 \text{ with probability } p$$

$$= \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} = A_2 \text{ with probability } 1 - p = q.$$

Note: $A_1 A_2 \neq A_2 A_1$ and

$$E[\Gamma] = \begin{bmatrix} a - \frac{a}{2}p & \frac{a}{2}p \\ \frac{a}{2}p & 1 - (1 - \frac{a}{2})p \end{bmatrix}. \tag{7}$$

Special Case: $p = \frac{1}{2}$, $C_{(0)}^T = [1,0]$, $a = e^{-\lambda t/m}$.

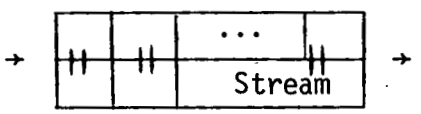
We can show that the limit as $m \to \infty$, $\Delta t \to 0$ such that $m\,\Delta t = t$,

$$\lim E[C_m] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \tfrac{1}{2} e^{-3\lambda t/4} \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{8}$$

Applications of this result to the retention of Strontium 85 (in dogs) is discussed in the paper by Uppuluri and Bernard (1967).

MAMMILLARY SYSTEMS

Next, the two-compartment model with a random valve, introduced by Uppuluri and Bernard, is extended to mammillary systems. (Each peripheral compartment communicates with the stream.)



In each case of a j-compartment mammillary system, we assume that during a small interval of time Δt, exactly one of the peripheral compartments is in communication with the stream. Symbolically we denote this by $(E_1, E_2, \ldots, E_{j-1})$, and if the i^{th} peripheral compartment is in communication with the stream, let $E_i = 1$ and $E_\ell = 0$ for $i \neq \ell$ (and $\ell = 1, 2, \ldots, j-1$). We may assign a probability p_i to this event, so that $p_1 + \ldots + p_{j-1} = 1$ and $E_1 + \ldots + E_{j-1} \equiv 1$. At any time $m\,\Delta t$, if the i^{th} peripheral compartment

is in communication with the stream, let us suppose that instantly the amount in the stream and the i^{th} peripheral compartment becomes

$$C_{s,m} = \frac{j-1}{j} a(C_{s,m-1} + C_{i,m-1})$$

$$C_{i,m} = \frac{1}{j} a(C_{s,m-1} + C_{i,m-1}) ,$$

and the amounts in the rest of the compartments remain the same as at time $(m-1)\Delta t$. This may be written as

$$\begin{bmatrix} C_{s,m} \\ C_{1,m} \\ \vdots \\ \\ C_{j-1,m} \end{bmatrix} = \begin{bmatrix} \frac{j-1}{j} a & \frac{j-1}{j} aE_1 & \frac{j-1}{j} aE_2 & \cdots & \frac{j-1}{j} aE_{j-1} \\ \frac{a}{j} E_1 & \frac{a}{j} E_1 + 1-E_1 & 0 & \cdots & 0 \\ \frac{a}{j} E_2 & 0 & \frac{a}{j} E_2 + 1-E_2 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{a}{j} E_{j-1} & 0 & 0 & & \frac{a}{j} E_{j-1} + 1-E_{j-1} \end{bmatrix} \begin{bmatrix} C_{s,m-1} \\ C_{1,m-1} \\ \\ \\ C_{j-1,m-1} \end{bmatrix}$$

or

$$C_m = \Gamma_m C_{m-1} . \tag{9}$$

Special Case $j = 3$: $E_1 = 1$, $E_2 = 0$ or $E_1 = 0$, $E_2 = 1$

$$\Gamma \in \left\{ A_1 = \begin{bmatrix} \frac{2a}{3} & \frac{2a}{3} & 0 \\ \frac{a}{3} & \frac{a}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ or } A_2 = \begin{bmatrix} \frac{2a}{3} & 0 & \frac{2a}{3} \\ 0 & 1 & 0 \\ \frac{a}{3} & 0 & \frac{a}{3} \end{bmatrix} \right\} \tag{10}$$

Further specializing to $a = 1$, we can get a complete description of the sample space of C_m. And as $m \to \infty$, we can obtain the limiting concentration vector to be $(\frac{1}{2},\frac{1}{4},\frac{1}{4})$; which is reasonable (when $a = 1$) in a three compartment system, where the volume of the stream is equal to the volume of the two peripheral compartments.

MOMENT GENERATING FUNCTIONS

Next, we obtain the functional equation satisfied by the moment generating function. Let $\tau' = (t_s, t_1, t_2, \ldots, t_{j-1})$, and $\phi_m(\tau')$ denote the moment generating function of the concentration vector C_m, and E denote the expectation operator.

$$\phi_m(\tau') = E \exp(\tau' C_m) = E \exp\left\{(\tau' \sum_{i=1}^{j-1} E_i A_i) C_{m-1}\right\}$$

$$= E \exp\left\{\sum_{i=1}^{j-1} (E_i \tau' A_i) C_{m-1}\right\} . \qquad (11)$$

Invoking the principle of independence, we obtain

$$\phi_m(\tau') = \sum_{i=1}^{j-1} p_i E \exp(\tau' A_i C_{m-1})$$

$$= \sum_{i=1}^{j-1} p_i \phi_{m-1}(\tau' A_i), \; m = 1,2,\ldots \qquad (12)$$

where

$\phi_0(\tau') = \exp(\tau' C_0)$ and C_0 is the vector of initial amounts in the j compartments.

FIRST MOMENTS

The 'mean' amount at the m^{th} stage is given by

$$\mu_m = E(C_m) = \sum_{i=1}^{m-1} p_i A_i \mu_{m-1} = M^m C_0, \quad m = 1,2,\ldots \qquad (13)$$

where

$$M = \sum_{i=1}^{j-1} p_i A_i . \qquad (14)$$

COVARIANCES

For the covariance (uncorrected) matrix Q_m of the amounts in the system at the m^{th} stage, we have

$$Q_m = E[C_m C_m'] = E[T\ C_{m-1}\ C_{m-1}' T']$$

$$= E[(\sum_{i-1}^{j-1} E_i A_i)\ C_{m-1}\ C_{m-1}'\ (\sum_{i=1}^{j-1} E_i A_i')]$$

$$= \sum_{i=1}^{j-1} p_i A_i Q_{m-1} A_i' , \quad m = 1,2... \tag{15}$$

where $Q_0 = C_0 C_0'$.

OTHER MODELS

Different models can be considered by changing the connectivity of the system, or the complex of communication, and/or the probability structure associated with the connecting valves. More details can be found in the paper by Bernard, Shenton and Uppuluri (1967).

APPLICATIONS IN DEMOGRAPHY AND CONTROL THEORY

Products of random matrices were used by Sykes (1967) and Pollard (1966) in stochastic population models. In order to compute the covariances, Pollard invoked the concept of the direct product of matrices. Grenander (1968) has shown how products of random matrices arise when one solves a second order linear differential equation

$$\frac{d^2x}{dt^2} + a^2(t)\ x(t) = G \tag{16}$$

where a(t) is a discrete stochastic process. In the study of the stability of random linear systems, Bharucha (1962) considered the linear differential system

$$\dot{x}(t) = A_k\ x(t), \text{ for } t_{k-1} \leq t < t_k,\ k = 1,.,... \tag{17}$$

where x(t) is an N x 1 vector, and A_k a random N x N matrix. For this model, the j^{th} moment convergence and almost sure convergence of the sequence $\{x(t_k)\}$ is studied; this involves products of random matrices.

CONSTANT PULSING AND RANDOM DECAY MODELS

In this section we study the random sequences defined by

$$Y_n = 1 + X_n\ Y_{n-1},\ n = 1,\ 2,\ ..., \tag{18}$$

with $Y_0 = 1$ and $\{X_n\}$ is a sequence of random variables. In a discrete time model, Y_n may refer to the amount of the substance at time $n\Delta t$; X_n denotes the random factor by which the amount at $(n-1)\ \Delta t$ is reduced during the interval of time Δt and 1 and is the constant amount which is added to the system during Δt. This may be referred to as a constant pulsing random decay model. It is of interest to know the weakest assumptions on $\{X_n\}$ such that the distributional properties and moments of Y_n converge. Such results are discussed in the paper by Uppuluri, Shenton and Feder (1967). We state below a few typical results without proofs.

Theorem 1: Let $Y_0 = 1$, $Y_n = 1 + X_n\, Y_{n-1}$, $n = 1, 2, \ldots,$ and $\{X_n\}$ be a sequence of independent identically distributed random variables.

Then

$$P[X_n = 1] = p = 1 - P[X_n = 0] \text{ if and only if}$$

$$P[Y_\infty = k] = (1-p)\, p^{k-1},\ k = 1, 2, \ldots \quad . \tag{19}$$

Theorem 2: Let $\{X_n\}$ be a sequence of independent identically distributed random variables, $Y_0 = 1$, and $Y_n = 1 + X_n\, Y_{n-1}$, $n = 1, 2, \ldots$.

Let

$$P[X_n = 1] = p,\ P[X_n = a] = 1 - p,\ 0 \leq a < 1,$$

$$\phi_n(t) = E[e^{tY_n}].$$

Then

$$\phi(t) = \lim_{n\to\infty} \phi_n(t) = \prod_{k=0}^{\infty} \frac{(1-p)e^{ta^k}}{1-pe^{ta^k}} \quad ,$$

and

$$Y_\infty = \sum_{k=0}^{\infty} a^k\, Z_k,$$

where Z_k are independent identically distributed random variables with

$$P[Z_k = m] = (1-p)\, p^m,\ m = 0, 1, 2, \ldots \quad .$$

Theorem 3: Under the hypothesis of theorem 2, we have the following type of central limit theorem:

$$\lim_{a\to 1^-} P\left[\frac{Y_\infty - E(Y_\infty)}{\sigma(Y_\infty)} \leq x\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

We have a more general

Theorem 4: Let $Y_n = 1 + X_n Y_{n-1}$, $Y_0 = 1$ and $\{X_n\}$ be a sequence of independent identically distributed non-negative random variables. Then $E(X_n) < 1$ implies that Y_n converges in distribution to a random variable Y_∞ and $E(Y_n)$ converges to $E(Y_\infty)$. Furthermore, if X is either a discrete or a continuous variable with $E(X^j) < 1$, $j = 1, 2, \ldots, m$, then $E(Y_n^j)$ converges to $E(Y_\infty^j)$, $j = 1, 2, \ldots, m$. In general, we have only convergence in distribution, but no convergence in probability of the sequence Y_n.

The limiting moments can be calculated by the recursive relationship

$$E(Y_\infty^j) = E[(1+X\ Y_\infty)^j], \tag{20}$$

the limit distribution function $F(x) = P[Y_\infty \leq x]$ satisfies the functional equation

$$F(x+1) = \int_0^\infty F(x/t)\, dG(t), \text{ where } G(t) = P[X_n \leq t]. \tag{21}$$

In a sense, our results establish the existence of unique solutions for integral equations of special type, by using probabilistic methods.

RANDOM PULSING AND CONSTANT DECAY MODELS

This model is defined by

$$Y_n = W_n + \rho\, Y_{n-1},\ n = 1, 2, \ldots,$$
$$Y_0 = 0, \text{ and } 1 > \rho > 0, \tag{22}$$

and $\{W_n\}$ is a sequence of independent identically distributed random variables. This is often used as an econometric model, and is referred to as an auto-regressive scheme. We used the recursive methods developed by us to study the behavior of this sequence. More details may be found in Uppuluri, Feder and Shenton (1967). We shall present here a few results.

Theorem 5: If $|E(W_n)| < \infty$, then Y_n converges in distribution to a random variable Y_∞ and $E(Y_n)$ converges to $E(Y_\infty)$. Furthermore, if $|E(W_n^j)| < \infty$, then $E(Y_n^j)$ converges to $E(Y_\infty^j)$, $j = 1, 2, \ldots, m$.

Theorem 6: If $Var(W_n) = \sigma^2 < \infty$, then $(1/r) \sum_{\alpha=1}^{r} Y_\alpha$ converges in mean square (as $r \to \infty$) to $E(W)/(1-\rho)$.

Theorem 7: If $E(W_n^4) < \infty$, then $(1/r) \sum_{\alpha=i+1}^{r} Y_\alpha Y_{r-\alpha}$ converges in mean square (as $r \to \infty$) to $E(W^2)\rho^i/(1-\rho)^2$.

Let us consider the special case

$\rho = 1/k$ and $P[W_n = j] = 1/k$; $j = 0, 1, 2, \ldots, (k-1)$. Then we have

Theorem 8: $Y_n = W_n = \rho W_{n-1} + \ldots + \rho^{n-1} W_1$ converges in distribution to Y_∞ where $P[Y_\infty \leq y] = y/k$, for $0 \leq y \leq k$.

Conversely, if $\rho = 1/k$, and Y_n converges in distribution to Y_∞ where

$P[Y_\infty \leq y] = y/k$ for $0 \leq y \leq k$, then

$W_n = 0, 1, 2, \ldots, (k-1)$, each with probability $1/k$.

RANDOM PULSING AND RANDOM DECAY MODELS

In this section, we consider the random difference equation

$$Y_n = U_n + V_n Y_{n-1}, \quad n = 1, 2, \ldots \tag{23}$$

$Y_0 = 1$ and $\{(U_n, V_n)\}$ is a sequence of mutually independent random variables with common distribution $G(u, v) = P[U_1 \leq u, V_1 \leq v]$, and U_n and V_n have marginal distributions $G_1(u)$ and $G_2(v)$ respectively. Let $F_n(y) = P[Y_n \leq y]$.

Such random difference equations arise in the study of mathematical models, where some substance is introduced into the system in random quantities, and the system periodically eliminates a random proportion of the substance available during a time interval. We are interested in the study of the behavior of the amount of the substance in the system after a "long" period of time. The mathematical aspects of this problem are considered in Paulson and Uppuluri (1972). We will present some results below.

Theorem 9: If for some $\alpha > 0$, $E|U_1|^{\alpha} < \infty$, $E|V_1|^{\alpha} < 1$, then there exists a distribution $F(y)$ such that $F_n(y)$ converges to $F(y)$ at every point of continuity of $F(\ )$.

Corollary: There exists a unique solution $\phi(t)$ of the characteristic functional equation

$$\phi(t) = E[e^{itU}\ \phi(Vt)] \tag{24}$$

where $\phi(t)$ is the characteristic function associated with the distribution $F(\ )$; and the expectation is relative to the measure induced by $G(u, v)$.

In the special case $\{U_n\}$ and $\{V_n\}$ are mutually independent sequences with common distribution functions $G_1(u)$ and $G_2(v)$ respectively; we have

$$\phi(t) = \Psi(t)\ E[\phi(Vt)] \tag{25}$$

has a unique solution $\phi(t)$, where $\Psi(t) = E[e^{itU_1}]$ and $E[\phi(Vt)]$ is with respect to the distribution $G_2(v)$.

In some special cases, we get some interesting mathematical results. For example, we have

Theorem 10: The solution $\phi(t)$ of the equation

$$\phi(t) = \Psi(t)\ E[\phi(Vt)]$$

is the characteristic function of the geometric distribution, if, and only if $\Psi(t)$ is the characteristic function of a geometric distribution and $G(v)$ is such that $P[V = 0] = p = 1 - P[V = 1]$ for $0 < p \leq 1$.

DIRECTIONS FOR FURTHER RESEARCH

Some of the results discussed earlier can be extended in several directions. Some of these results were extended to the vector case by Paulson and Uppuluri. The distributions of the pulsing mechanisms can be parametrized, and methods can be developed to do statistical inference on these parameters. Some of the results could be extended to higher order random difference equations.

APPLICATION OF GENETICS

Random difference equations with additive and multiplicative structures were set up as mathematical models for the study of cultural inheritance by

Cavalli-Sforza and Feldman (1973). Only the behavior of the first two moments in such sequences were used in the interpretations. There is a need to study the behavior of these sequences in the context of cultural inheritance.

APPLICATIONS TO RECORD VALUES

Certain mathematical aspects of the sequences with additive and multiplicative structures were also studied recently by Vervaat (1974), in the context of record values and inter-record times. Vervaat gives certain sufficient conditions in order that the sequence may converge in distribution. There is a need to study the qualitative nature of these sufficient conditions and those of Paulson and Uppuluri (1972), and attempt to find necessary and sufficient conditions for the convergence of these sequences. In the same spirit one should study Kesten (1973) and Solomon (1975).

MORE GENERAL SCHEMES AND ISING MODELS

Bellman, Soong and Vasudevan (1972) studied the asymptotic behavior of sequences of the form

$$x_{n+1} = g(x_n, r_n) \tag{26}$$

where x_n is a N-dimensional state vector and the vector sequence $\{r_n\}$ is assumed to constitute a time homogeneous simple Markov process. As an application of their approach, they studied the behavior of the partition function Q_N of an Ising model as the number of spins N becomes large. They indicated that their approach is promising to the study of higher dimensional Ising models and said that they would discuss the results elsewhere. Indeed, it will be very exciting to see the techniques and results of random difference equations used in getting a better understanding of the higher dimensional Ising models.

RANDOM MATRICES

As a matter of fact, in the one dimensional Ising model with random impurities, the results of Furstenberg (1963) were used by McCoy and Wu (1973) to compute the free energy (in the thermodynamic limit) in the absence of

a magnetic field. McCoy and Wu indicated that all aspects of the random Ising model have not been carefully studied.

In 1970, Hammersley and Menon (1970) discussed the classical unsolved problem of solid-state chemistry, which goes under the name of the monomer-dimer problem. While discussing about random determinants, on page 348, they say:

> "Some sort of limit theory for random matrices or random determinants, and especially products of random matrices, would be useful here; but the literature (Mehta [15]) seems virtually silent on the subject; the nearest approach (Furstenberg and Kesten [7]) is not applicable since our random matrix contains zeros; and, even if it were applicable, it would only assert the existence of a limit, rather than help in evaluating the limit."

There is a need to study the monomer-dimer problem in view of some of the recent work on products of random matrices.

For a beautiful expository article on random matrices in physics, refer to Wigner (1967). At the end of the article, he enumerates a number of general questions and hopes that the mathematicians will help to clear up some of them. For the mathematical aspects of random operators, refer to the survey article by Pastur (1973).

RANDOM FIXED POINT THEOREMS

While studying the ergodic behavior of the log norms of products of random matrices, Grenander (1968), on page 163, invokes a stochastic version of the Banach Fixed Point Theorem. Kalman (1962) invokes the Brouwer fixed point theorem in order to establish the convergence of the second moments of products of random matrices. Kesten (1973) invokes the Shauder-Tychonoff fixed point theorem, while establishing a result about the limiting behavior of the norm of products of random matrices.

The theory of random fixed point theorems is not well developed. The best known result seems to be due to Krasnoselskii, mentioned by Bharucha-Reid (1972). It should be very illuminating to prove the existence of solutions of functional equations that arose in the study of random sequences of this paper, through the corresponding random fixed point theorems.

REFERENCES

Bellman, R., Soong, T. T., and Vasudevan, R., "On Moment Behavior of a Class of Stochastic Difference Equations," J. Math. Analys. and its Appl., 40 (1972), 286-299.

Bernard, S. R., Shenton, L. R. and Uppuluri, V. R. R., "Stochastic Models for the Distribution of Radioactive Material in a Connected System of Compartments," Proceedings of the Fifth Berkeley Symposium, IV (1967), 481-510.

Bharucha, B. H., "Stability of Random Linear Systems with Markov Parameters," ERD-TN-60-776, Scientific Report No. 2, Electronics Research Lab., Dept. of Electrical Engineering, University of California, Berkeley, 1962.

Bharucha-Reid, A. T., Random Integral Equations, Academic Press, New York and London, 1972.

Cavalli-Sforza, L. and Feldman, M. W., "Models for Cultural Inheritance I. Group Mean and Within Group Variation," Theoretical Population Biol., 4 (1973), 42-55.

Furstenberg, H., "Non-commuting Random Products," Trans. Am. Math. Soc., 108 (1963), 377-428.

Grenander, U., Probabilities on Algebraic Structures, 2nd ed., Almquist and Wiksell, 1968.

Hammersley, J. M. and Menon, V. V., "A Lower Bound for the Monomer-Dimer Problem," J. Inst. Maths. Applics, 6 (1970), 341-364.

Jacquez, J. A., Compartmental Analysis in Biology and Medicine, Elsevier Publishing Company, Amsterdam, New York, 1972.

Kalman, R. E., "Control of Randomly Varying Linear Dynamical Systems," Proc. Symp. in Appl. Maths., Am. Math. Soc., Providence, Rhode Island, 13 (1962), 287-298.

Kesten, H., "Random Difference Equations and Renewal Theory for Products of Random Matrices," Acta Mathematica, 131 (1973), 207-248.

McCoy, B. M. and Wu, Tai Tsun, The Two-Dimensional Ising Model, Harvard University Press, 1973.

Pastur, L. A., "Spectra of Random Self Adjoint Operators," Russian Mathematical Surveys, (1973), 1-65.

Paulson, A. S. and Uppuluri, V. R. R., "Limit Laws of a Sequence Determined by a Random Difference Equation Governing a One-Compartment System," Math. Biosciences, 13 (1972), 325-333.

Pollard, J. H., "On the Use of Direct Matrix Product in Analyzing Certain Stochastic Population Models, Biometrika, 53 (1966) 397-415.

Solomon, F., "Random Walks in Random Environment," Ann. Prob., 3 (1975), 1-31.

Sykes, Z. M., "A Stochastic Version of the Matrix Model for Population Dynamics," Proceedings of the 36th Session of the International Statistical Institute, Sydney, Australia, 1967.

Uppuluri, V. R. R. and Bernard, S. R., "A Two-compartment Model with a Random Connection," Compartments, Pools and Spaces in Medical Physiology, ed. Per-Erik Bergner and C. C. Lushbaugh, U.S. AEC Division of Technical Information, (1967), 175-188.

Uppuluri, V. R. R., Feder, P. I. and Shenton, L. R., "Random Difference Equations Occurring in One-compartment Models," Math. Biosciences, 1 (1967), 143-171.

Vervaat, Wim, "On Records, Maxima and a Stochastic Difference Equation," Report 73-11, Department of Mathematics, University of Amsterdam, The Netherlands, 1974.

Wigner, E. P., "Random Matrices in Physics," SIAM Review, 9 (1967), 1-23.

SIMULATION BOUNDS FOR SYSTEM AVAILABILITY

G.L. Tietjen
R.A. Waller
Los Alamos Scientific Laboratory
Los Alamos, New Mexico

System availability is a dominant factor in the practicality of nuclear power electrical generating plants. A proposed model for obtaining either lower bounds or interval estimates on availability uses observed data on n failure-to-repair cycles of the system to estimate the parameters in the time-to-failure and time-to-repair models. These estimates are then used in simulating failure/repair cycles of the system. The availability estimate is obtained for each of 5000 samples of n failure/repair cycles to form a distribution of estimates. Specific percentile points of those simulated distributions are selected as lower simulation bounds or simulation interval bounds for the system availability. The method is illustrated with operational data from two nuclear plants for which an exponential time-to-failure and a lognormal time-to-repair are assumed.

INTRODUCTION AND DEFINITIONS

A useful measure of the performance of a large system is availability. We define the availability of a system to be the probability of successful operation. In that sense we may speak of either "successful operation at some time t" or "successful operation over a long interval of t." We refer to the availability at time t as a point or transient availability; whereas, the availability over a long operating period is called steady-state or inherent or long-run availability. This paper presents a method which may be used to generate simulation bounds for either point or steady-state availability under specific distributional assumptions. However, the examples presented are restricted to a study of the steady-state availability alone.

For illustration we suppose a nuclear power plant is placed in service and that its performance is followed through n failure/repair cycles. We use $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ to denote the observed times-to-failure and times-to-repair during the observation period. Further, we assume that the X's and Y's are mutually independent random variables such that X_i, i=1, 2, ..., n has an exponential density

$$f(x) = \lambda\exp(-\lambda x),\ x > 0 \tag{1}$$

and that Y_i, i=1, 2, ..., n has a lognormal density

$$h(y) = (\sqrt{2\pi}\ \beta y)^{-1} \exp\left\{(-1/2)\left[(\ell n\ y - \alpha)/\beta\right]^2\right\},\ y > 0 \tag{2}$$

STEADY-STATE AVAILABILITY

Using the definition of steady-state availability given by Gray and Lewis (1967), Gray and Schucany (1969), Nelson (1970), and others, we have

$$A = \mu_x/(\mu_x+\mu_y) = \lambda^{-1}/[\lambda^{-1} + \exp(\alpha+\beta^2/2)] = [1 + \lambda\exp(\alpha+\beta^2/2)]^{-1} \tag{3}$$

in which $\mu_x = E(X)$ is the mean time-to-failure and $\mu_y = E(Y)$ is the mean time-to-repair, Gray and Lewis (1967) and Gray and Schucany (1969) have presented methods for obtaining confidence intervals and bounds for A, but their work assumes that the variance β^2 is known. To avoid the restriction of known β^2 we present an ad hoc methodology for using observed data and computer simulation to derive simulation intervals and bounds for A. The simulation results for A are compared with confidence intervals provided by the Gray and Lewis (1967) method to study the usefulness of the proposed methodology.

METHODOLOGY FOR SIMULATION BOUNDS

We notice that A is a function of the three parameters λ, α, and β. Thus, we can obtain an estimator of A, say $\hat{A}$, by substituting estimators of λ, α, and β, say $\hat{\lambda}$, $\hat{\alpha}$, and $\hat{\beta}$, into equation (3). The maximum likelihood estimators are:

$$\hat{\lambda} = n/\sum_1^n x_i,$$

$$\hat{\alpha} = (1/n)\sum_1^n \ell n\ y_i, \tag{4}$$

$$\hat{\beta}^2 = (1/n)\sum_1^n (\ell n\ y_i - \hat{\alpha})^2$$

By substituting those estimators into equation (3), we obtain the maximum likelihood estimator of A.

To develop simulation intervals for A we proceed as follows:

Step 1: Use the observed times-to-failure, $X_1, X_2, \ldots, X_n$, and the observed times-to-repair, $Y_1, Y_2, \ldots, Y_n$, to calculate $\hat{\lambda}$. $\hat{\alpha}$ and $\hat{\beta}^2$ as defined in equation (4).

Step 2: Use $\hat{\lambda}$ as λ in equation (1) to generate a simulated sample of n exponential times-to-failure, $X_{s1}, X_{s2}, \ldots, X_{sn}$. Use $\hat{\alpha}$ and $\hat{\beta}^2$ as α and β^2 in equation (2) to generate a simulated sample of n lognormal times-to-repair, $Y_{s1}, Y_{s2}, \ldots, Y_{sn}$.

Step 3: Use the simulated data in Step 2 to evaluate parameter estimators, $\hat{\lambda}_s$, $\hat{\alpha}_s$, and $\hat{\beta}_s^2$ by equation (4). The subscript s indicates simulated data.

Step 4: Evaluate

$$\hat{A} = [1 + \hat{\lambda}_s \exp(\hat{\alpha}_s + \hat{\beta}_s^2/2)]^{-1}.$$

Step 5: Repeat Steps 2, 3 and 4 5000 times to simulate a distribution of $\hat{A}$.

Step 6: Determine the $(1 - \gamma)$% simulation intervals for A by using the $(\gamma/2)$ and $(1 - \gamma/2)$ percentile values of the simulated distribution of $\hat{A}$.

The examples in the next section illustrate the proposed methodology.

EXAMPLES

The United States Atomic Energy Commission document WASH 1203-73 (1973) provides graphical histories for the performance of nuclear-powered electrical generating plants operating in the United States. The data used in the two following examples are the histories of two such power plants. In the examples, x denotes the i^{th} time-to-failure and y denotes the i^{th} time-to-repair.

EXAMPLE 1

The data in Table 1 are the times-to-failure and times-to-repair measured in years for n=19 failure/repair cycles observed for the Yankee Nuclear Power Plant in Rowe, Massachusetts during operation from December 1960 to May 1973. For these data n=19, $\sum x_i = 9.603$, $\sum \ell n\ y_i = -52.744$, and $\sum(\ell n\ y_i - \hat{\alpha})^2 = 18.251$. Therefore,

TABLE 1. Times-to-Failure (x_i) and Times-to-Repair (y_i) for 19 Failure/Repair Cycles of the Yankee Nuclear Power Plant

i	x_i	y_i
1	0.063	0.027
2	0.055	0.038
3	0.296	0.014
4	0.170	0.036
5	0.822	0.345
6	0.948	0.197
7	0.715	0.096
8	0.923	0.255
9	0.899	0.090
10	0.332	0.033
11	0.304	0.049
12	0.658	0.107
13	0.523	0.019
14	0.712	0.148
15	0.485	0.022
16	0.397	0.030
17	0.145	0.101
18	0.912	0.019
19	0.244	0.260
Total	9.603	1.886

$$\hat{\lambda} = 19/9.603 = 1.979$$
$$\hat{\alpha} = -52.744/19 = -2.776 \qquad (5)$$
$$\hat{\beta}^2 = 18.251/19 = 0.961$$

Therefore, the maximum likelihood estimate of A as given by the evaluation of equation (3) is

$$\hat{A} = (1 + 1.979\exp(-2.776 + 0.961/2)]^{-1} = 0.834$$

For comparative purposes we can consider another estimator of A to be the ratio of the "successful operating time" to "total time." That is,

$$\tilde{A} = \sum_{i=1}^{n} x_i \Big/ \sum_{i=1}^{n} (x_i + y_i) = 9.603/11.489 = 0.836 \qquad (6)$$

We observe that the estimates $\hat{A}$ and $\tilde{A}$ are very nearly equal.

Some selected percentile values for the simulated distributions of $\hat{A}$ and $\tilde{A}$ as given in Table 2 were determined by the simulation method outlined earlier.

TABLE 2. Simulation Percentile Values for $\hat{A}$ and $\tilde{A}$

Statistic	P = 0.025	0.050	0.500	0.950	0.975
$\hat{A}$	0.710	0.734	0.834	0.899	0.909
$\tilde{A}$	0.711	0.734	0.836	0.901	0.910

From Table 2 we find that 90% simulation interval for A is either

90% SI = $0.734 \leqq A \leqq 0.899$ (based on $\hat{A}$)

or 90% SI = $0.734 \leqq A \leqq 0.901$ (based on $\tilde{A}$).

For comparison, we can assume that $\beta^2 = 0.961$ is a known value (rather than estimated as in equation (5)) and use interpolation and extrapolation for the tables in Gray and Lewis (1967) to find that a 90% CI for A is

$$0.748 \leqq A \leqq 0.894.$$

Thus, there is close agreement between the simulation bounds and the confidence intervals based on the estimated variance, $\hat{\beta}^2$.

We can use Table 2 to set other level simulation intervals and/or lower bounds for A.

EXAMPLE 2

The data in Table 3 are the times-to-failure and times-to-repair measured in years for 18 failure/repair cycles observed for the operation of the Humboldt Bay Power Plant, Unit No. 3, Eureka, California, from October 1963 to December 1974. For these data n = 18, $\sum x_i = 10.055$, $\sum \ell n\ y_i = -49.729$, and $\sum(\ell n\ y_i - \hat{\alpha})^2 = 10.496$. Therefore,

$$\hat{\lambda} = 18/10.055 = 1.790$$
$$\hat{\alpha} = -49.729/18 = -2.763 \qquad (7)$$
$$\hat{\beta}^2 = 10.496/18 = 0.583$$

The maximum-likelihood steady-state availability estimate as given by the evaluation of equation (3) is

$$\hat{A} = [1 + 1.790\ \exp(-2.763 + 0.583/2)]^{-1} = 0.869.$$

Further,

$$\tilde{A} = 10.055/11.526 = 0.872.$$

Again, the values of $\hat{A}$ and $\tilde{A}$ are in close agreement.

Table 4 presents some selected percentile values for the simulated distributions of $\hat{A}$ and $\tilde{A}$.

TABLE 3. Times-to-Failure (x_i) and Times-to-Repair (y_i) for 18 Failure/Repair Cycles of the Humboldt Bay Power Plant, Unit No. 3

i	x_i	y_i
1	0.523	0.060
2	0.175	0.038
3	0.537	0.074
4	1.019	0.197
5	0.121	0.016
6	0.827	0.088
7	0.271	0.016
8	0.499	0.066
9	0.940	0.058
10	0.466	0.099
11	0.742	0.060
12	0.189	0.058
13	0.422	0.016
14	0.389	0.222
15	1.000	0.118
16	0.003	0.047
17	0.855	0.085
18	1.077	0.153
Total	10.055	1.471

TABLE 4. Percentile Values for Simulated Distributions of $\hat{A}$ and $\tilde{A}$

Statistic	P = 0.025	0.050	0.500	0.950	0.975
$\hat{A}$	0.779	0.796	0.868	0.915	0.923
$\tilde{A}$	0.780	0.797	0.868	0.915	0.923

From Table 4 we find

$$90\%\ SI = 0.796 \leq A \leq 0.915 \qquad \text{(based on } \hat{A}\text{)}$$

and $$90\%\ SI = 0.797 \leq A \leq 0.915 \qquad \text{(based on } \tilde{A}\text{)}.$$

By using $\beta^2 = 0.583$ as a known value the Gray and Lewis (1967) confidence interval procedure gives (with interpolation and extrapolation)

$$90\%\ CI = 0.778 \leq A \leq 0.901.$$

EVALUATION OF THE SIMULATION PROCEDURE

To study the sampling behavior of the simulation procedure and to investigate a "confidence" level for simulation intervals, we assumed that λ, α, and β^2 were known to be equal to the estimates provided by the data in the preceding examples. For Example 1, we set $\lambda = 1.979$, $\alpha = -2.776$ and $\beta^2 = 0.961$. Then the true availability is $A = 0.834$. Next we generated 1000 samples of nineteen failure/repair cycles. Each of these 1000 samples provided estimates of λ, α, and β^2 which were used to determine a 95% simulation interval for A. The interval was determined by the method outlined earlier except that each interval was based on 100 rather than 5000 estimates of A. The results showed that 94.2% of the simulation intervals based on $\hat{A}$ and 94% of the simulation intervals based on $\tilde{A}$ included the true value of $A = 0.834$. As a second study we followed the above plan to generate 100 95% simulation intervals where each interval was based on 1000 simulated samples. Here we observed that 97% of the 95%-simulation intervals for both $\hat{A}$ and $\tilde{A}$ included $A = 0.834$.

For Example 2, we repeated the above study based on 18 failure/repair cycles, when $\lambda = 1.790$, $\alpha = -2.763$, $\beta^2 = 0.583$. The "confidences" based on 1000 samples of 100 95%-simulation intervals were 95% for $\hat{A}$ and 94.5% for $\tilde{A}$. Similarly, the "confidences" based on 100 samples of 1000 95%-simulation intervals were 95% for both $\hat{A}$ and $\tilde{A}$.

Thus, it seems that the simulation intervals are exhibiting confidence levels which are nearly equal to the difference in percentiles used in their computation.

CONCLUSIONS

The simulation procedure proposed here is simple to use and requires about 10 seconds of computer time (CDC 6600) to provide any desired percentiles of the simulated distribution of $\hat{A}$. Further, any brief investigation of the confidence level of the procedure indicates that it performs very well. These facts along with the fact that we do not need to assume knowledge of any parameters in the models provide strong support for the procedure.

The proposed procedure may be used to find intervals or bounds for transient availability so long as an expression for A(t) is available. The development in Chapter 7 (p. 192) of Barlow and Proschan (1975) can be used to obtain A(t). Subsequent studies are planned for this area of availability estimation.

REFERENCES

Barlow, R. E. and Proschan, F., Statistical Theory of Reliability and and Life Testing Probability Models, Holt, Rinehart and Winston, Inc., New York, (1975), 192.

Gray, H. L. and Lewis, T. O., "A Confidence Interval for the Availability Ratio." Technometrics, 9, No. 3, (1967), 465-471.

Gray, H. L. and Schucany, W. R., "Lower Confidence Limits for Availability Assuming Lognormally Distributed Repair Times." IEEE Transactions on Reliability, R-18, No. 4, (1969), 157-162.

Nelson, W., "A Statistical Prediction Interval for Availability." IEEE Transactions on Reliability, R-19, No. 4, (1970), 179-182.

U.S. Atomic Energy Commission, Operating History: U.S. Nuclear Power Reactors, WASH 1203-73, Division of Reactor Development and Technology, 1973.

RADIOCHEMICAL SPECTRAL ANALYSIS BY MAXIMUM LIKELIHOOD

W. L. Nicholson
D. L. Stevens, Jr.
Battelle Pacific Northwest Laboratories
Richland, Washington

Maximum likelihood estimation provides a superior alternative to weighted least squares for low count situations. This approach extracts information from both sample and background to estimate background level in each energy bin. Confidence intervals for individual nuclide estimates can be constructed from likelihood contours. If the precision of the standard spectra is of the same order as that of the gross sample, maximum likelihood can model the randomness in standards and include this additional variability in standard deviation estimates and confidence intervals for individual radionuclides.

INTRODUCTION

The fitting of a linear combination of energy spectral shapes for radioactive decay of pure radionuclides to a net sample energy spectrum is a classical method in quantitative radiochemistry. Weighted least squares is the usual estimation procedure for resolution of the pure nuclides in the sample mixture. Many researchers use this technique to resolve mixtures of gamma-ray emitting radionuclides from pulse-height distributions collected in either sodium iodide or lithium-drifted germanium detectors. While not extensively reported in the literature, the technique is equally valid for any energy spectrum recorded as a frequency histogram of radioactivity-generated events versus energy. An example is isotopic resolution of noble gases by liquid scintillation spectroscopy as done by Horrocks and Studier (1964). Here, alpha particles, conversion electrons and beta particles, as well as gamma rays, can contribute to the spectrum. Turkevich, et al., (1967) determine the chemical composition of the lunar surface at the Surveyor V site by weighted least squares analysis of backward-scattered alpha particle and proton spectra. The common ingredient of such radiochemical experiments is that the total number of events in a specified energy

range during a fixed time interval follows the Poisson distribution probability law.* Further, since each event is sensed by a detection system with a constant efficiency, the number of counts, or recorded events in a fixed time period, also follows a Poisson distribution probability law.

In a sample spectrum of many energy bins, the number of counts per energy bin may vary over several orders of magnitude. The variance of the Poisson distribution is equal to the mean. Hence, by the Gauss-Markov theorem (Rao, 1965), any near efficient least squares procedure must weight individual net counts as approximately inversely proportional to their variance, which is a linear combination of gross sample and background means. In practice, weights are estimated directly from gross sample and background spectra counts. Slavinskas, et al., (1966) consider alternative ways of determining weights. Only for low count situations are results highly dependent on the weighting scheme. When all counts are reasonably large, inferences can be based on asymptotic normal distribution theory. Pasternack and Liuzzi (1966) show that this asymptotic theory is good enough for residual analysis and use chi-square and von Neumann ratio (1941) statistics to measure lack of fit.

With few counts in an energy bin of gross sample and background spectra the weight is poorly determined. This situation is typical of low-level radiochemical analysis at the threshold of detection. Now weighted least squares estimation may give anomalous results. Weights may emphasize the wrong portion of the spectrum. Also, with incorrect weights, standard deviation estimates are poor. Maximum likelihood estimation is an attractive alternative for low count situations. Even for a zero or one count, maximum likelihood extracts the information in gross sample and background spectra. In the weighted least squares analysis, the background spectrum alone is used to estimate background and a net sample spectrum is obtained by subtraction. The maximum likelihood approach uses the information in both sample and background to estimate the background level in each energy bin. Confidence intervals for individual nuclide estimates can be constructed from likelihood contours. This is superior to the weighted least squares

*If the event is radioactive decay, the half-life of the nuclide is assumed long compared to the length of the time period of observation, for otherwise the correct probability law would be the binomial.

asymptotic normal theory approach. The non-negativeness of each nuclide constraint is handled naturally by maximization over the feasible set of parameter values (i.e., all components non-negative) with likelihood contours similarly constrained.

In most practical situations the standard spectra are known to a much greater precision than either sample or background. However, if the precision of the standards is of the same order as that of the gross sample, maximum likelihood can model the randomness in standards and include this additional variability in standard deviation estimates and confidence intervals for individual radionuclides.

In the next section the radiochemical spectral analysis problem is expressed mathematically. Maximum likelihood estimates are derived for the two situations of standard spectra well-known and standard spectra imprecise. Next, the maximum likelihood method is applied to the separation of radioiodine isotopes by neutron activation. Here, both sample and standards are very low level. In the maximum likelihood solutions, the standards are assumed to be estimates.

DERIVATION OF MAXIMUM LIKELIHOOD ESTIMATES

Let Y_i denote the gross sample counts in the i^{th} energy bin, $i = 1, 2, \ldots, q$, and let $Y = (Y_1, Y_2, \ldots, Y_q)^T$. The mean of Y is assumed to be a linear combination of a background spectrum γ and p standard spectra $x_1, x_2, \ldots, x_p$:

$$E[Y] = \gamma + \sum_j x_j \beta_j$$

$$= \gamma + X\beta$$

where

$$\gamma = (\gamma_1, \ldots, \gamma_q)^T$$

$$x_j = (x_{1j}, \ldots, x_{qj})^T$$

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{q1} & \cdots & x_{qp} \end{pmatrix}$$

and

$$\beta = (\beta_1, \ldots, \beta_p)^T.$$

Thus, $\beta_j \sum_i x_{ij}$ is the total count attributed to the j^{th} nuclide.

Let Z_i denote the background counts in the i^{th} energy bin, and let K_o be the ratio of the background counting time to the sample counting time, so that

$$E[Z_i] = K_o\gamma_i$$

or, for $Z = (Z_1, \ldots, Z_q)^T$

$$E[Z] = K_o\gamma.$$

STANDARD SPECTRA KNOWN

In most practical situations the standard spectra are determined to a much greater precision than either sample or background. If this is the case, the X matrix is treated as known. The situation in which the X matrix is also to be estimated will be discussed later.

The variables Y_i and Z_i are mutually independent Poisson variables, so the log likelihood function is

$$\ell(Y,Z|\beta,\gamma) = \sum_q [Z_i \ell n(K_o\gamma_i) + Y_i \ell n(x^i\beta+\gamma_i) - (K_o+1)\gamma_i - x^i\beta] + F(Y,Z)$$

where $x^i = (x_{i1}, x_{i2}, \ldots, x_{ip})$.

Taking partial derivatives and equating to zero yields the likelihood equations:

$$\frac{\partial \ell}{\partial \gamma_i} = \frac{Z_i}{\gamma_i} + \frac{Y_i}{x^i\beta+\gamma_i} - (K_o+1) = 0 \qquad (1)$$

$$l = 1, 2, \ldots, q$$

$$\frac{\partial \ell}{\partial \beta_j} = \sum_q \left(\frac{Y_i x_{ij}}{x^i\beta+\gamma_i} - x_{ij} \right) = 0 \qquad (2)$$

$$j = 1, 2, \ldots, p$$

By defining

$$W = \mathrm{diag}(x^1\beta+\gamma_1,\ x^2\beta+\gamma_2,\ \ldots,\ x^q\beta+\gamma_q) \tag{3}$$

equations (2) can be written as

$$X^TW^{-1}(Y-\gamma) = X^TW^{-1}X\beta. \tag{4}$$

Equations (4) are in the form of the normal equations for weighted least squares, with the exception that the weighting matrix W is a function of the parameter vector β.

By clearing equations (1) of fractions and collecting terms with like powers of γ_i:

$$(K_o+1)\gamma_i^2 + [(K_o+1)x^i\beta - Y_i - Z_i]\gamma_i - Z_ix^i\beta = 0 \tag{5}$$

$$i = 1, 2, \ldots, q.$$

These can be solved for γ_i, conditional on an estimate of β, by using the quadratic formula. Asymptotic theory indicates that the positive radical should be used.

Equations (3), (4), and (5) constitute the basis for an algorithm for the solution to the likelihood equations. The algorithm has the form

$$\tilde{Y}_i = \max(Y_i, 1)$$

$$\tilde{W}(0) = \mathrm{diag}(\tilde{Y}_i, \ldots, \tilde{Y}_q)$$

$$\tilde{\gamma}(0) = (1/K_o\ Z)$$

$$\tilde{\beta}(k) = [X^T\tilde{W}(k-1)^{-1}X]^{-1}X^T\tilde{W}(k-1)^{-1}[Y-\tilde{\gamma}(k-1)];$$

$$\tilde{\gamma}_i(k) = \frac{Y_i + Z_i - (K_o+1)x^i\tilde{\beta}(k) + \left\{[Y_i+Z_i-(K_o+1)x^i\tilde{\beta}(k)]^2 + 4(K_o+1)Z_ix^i\tilde{\beta}(k)\right\}^{1/2}}{2(K_o+1)}$$

$$i = 1, 2, \ldots, q;$$

$$\tilde{W}(k) = \mathrm{diag}[x^1\tilde{\beta}(k) + \tilde{\gamma}_1(k), \ldots, x^q\tilde{\beta}(k) + \tilde{\gamma}_q(k)],$$

for $k = 1, 2, \ldots$.

Then the maximum likelihood estimates are

$$\hat{\gamma} = \lim_{k\to\infty} \tilde{\gamma}(k)$$

and

$$\hat{\beta} = \lim_{k\to\infty} \tilde{\beta}(k).$$

Algorithms of this form are known to converge provided the initial estimates are "reasonably" accurate and the functions involved are "reasonably" well behaved. A convergence proof for this particular problem is difficult because of the number of parameters involved; however, in all cases tried, convergence has been rapid.

For θ_k, k = 1, 2, a vector of parameters, define

$$S(\theta_1,\theta_2) = E\left[\left(\frac{\partial \ell}{\partial \theta_1}\right)\left(\frac{\partial \ell}{\partial \theta_2}\right)^T\right].$$

Then from maximum likelihood theory (e.g., Wilks, 1962, p. 380) the asymptotic covariance matrix of the estimates is

$$V = \mathrm{Cov}\begin{pmatrix}\hat{\beta}\\ \hat{\gamma}\end{pmatrix} = \left(\begin{array}{c|c} S(\beta,\beta) & S(\beta,\gamma) \\ \hline S(\gamma,\beta) & S(\gamma,\gamma)\end{array}\right)^{-1}$$

$$= \left(\begin{array}{c|c} V_{\beta\beta} & V_{\beta\gamma} \\ \hline V_{\gamma\beta} & V_{\gamma\gamma}\end{array}\right).$$

Since the primary interest is in the estimated coefficients $\hat{\beta}$ and their associated covariance $V_{\beta\beta}$, only the matrix $V_{\beta\beta}$ is found. By partitioning V^{-1} as indicated, it follows that

$$V_{\beta\beta} = [S(\beta,\beta) - S(\beta,\gamma)S(\gamma,\gamma)^{-1}S(\gamma,\beta)]^{-1}.$$

Define

$$\mu_i = E[Y_i] = \gamma_i + x^i\beta$$

$$M = \mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_q)$$

and $$\Gamma = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_q).$$

Then, from the Poisson structure of Y and Z,

$$S(\beta,\beta) = X^T M^{-1} X$$

$$S(\beta,\gamma) = X^T M^{-1} = S(\gamma,\beta)^T$$

$$S(\gamma,\gamma) = M^{-1} + K_o\Gamma^{-1}$$

and $$V_{\beta\beta} = [X^T M^{-1} X - X^T M^{-1}(M^{-1}+K_o\Gamma^{-1})^{-1} M^{-1}X]^{-1}$$
$$= [X^T K_o(K_o M+\Gamma)^{-1}X]^{-1}$$
$$= (X^T W_o^{-1}X)^{-1}$$

where $W_o = M + (1/K_o)\Gamma$.

The maximum likelihood estimate of the asymptotic covariance $V_{\beta\beta}$ is

$$\hat{V}_{\beta\beta} = (X^T \hat{W}_o X)^{-1} \text{ with}$$
$$\hat{W}_o = \hat{M} + (1/K_o)\hat{\Gamma}.$$

The "^" notation $\hat{M}$ and $\hat{\Gamma}$ indicates substitution of $\hat{\gamma}_i$ and $\hat{\beta}$ in the definition of μ_i, M and Γ.

STANDARD SPECTRA TO BE ESTIMATED

If the precision of the standard spectra is approximately the same as that of the gross sample, this additional variability should be included in standard deviation estimates. Let Y and Z be as before, and let S_{ij} be the gross count in the i^{th} energy bin for the j^{th} standard, i = 1, ..., q; j = 1, ..., p. Then

$$E[S_{ij}] = K_j[x_{ij} + \gamma_i],$$

where K_j is the ratio of counting time for the j^{th} standard to the sample counting time. Let

$$S = \begin{bmatrix} S_{11} & \cdots & S_{1p} \\ \vdots & & \vdots \\ S_{q1} & \cdots & S_{qp} \end{bmatrix}.$$

The log likelihood function is

$$\ell(Y,Z,S|\beta,\gamma, S)$$
$$= \sum_i \left\{Y_i \ell n(x^i\beta+\gamma_i) + \sum_j S_{ij} \ell n[K_j(x_{ij}+\gamma_i)] + Z_i \ell n(K_o\gamma_i)\right.$$
$$\left. - x^i\beta - \sum_j x_{ij} - (\sum_j K_j+K_o+1)\gamma_i\right\} + F(Y,Z,S)$$

and the likelihood equations are

$$\frac{\partial \ell}{\partial \beta_j} = \sum_i \left\{ \frac{Y_i x_{ij}}{x^i\beta+\gamma_i} - x_{ij} \right\} = 0, \qquad j = 1, 2, \ldots, p; \tag{6}$$

$$\frac{\partial \ell}{\partial \gamma_i} = \frac{Z_i}{\gamma_i} + \frac{Y_i}{x^i\beta+\gamma_i} + \sum_j \frac{S_{ij}}{x_{ij}+\gamma_i} - \sum_j K_j - 1 - K_o = 0, \tag{7}$$

$$i = 1, 2, \ldots, q;$$

$$\frac{\partial \ell}{\partial x_{ij}} = \frac{Y_i \beta_j}{x^i\beta+\gamma_i} + \frac{S_{ij}}{x_{ij}+\gamma_i} - (\beta_j + K_j) = 0, \tag{8}$$

$$i = 1, 2, \ldots, q,$$

$$j = 1, 2, \ldots, p.$$

Equations (6) are the same as equations (3) and can be written in the form of normal equations:

$$X^T W^{-1}(Y-\gamma) = X^T W^{-1} X\beta.$$

By summing equations (8) over j and subtracting equations (7), there results a set of quadratic equations in γ_i:

$$\sum_j \frac{\partial \ell}{\partial x_{ij}} - \frac{\partial \ell}{\partial \gamma_i} = K_o + 1 - \sum_j \beta_j - \frac{Z_i}{\gamma_i} + \frac{(\sum_j \beta_j - 1)Y_i}{x^i\beta+\gamma_i} = 0 \tag{10}$$

$$i = 1, 2, \ldots, q$$

or

$$(K_o+1-\sum_j \beta_j)\gamma_i^2 + [(K_o+1-\sum_j \beta_j)x^i\beta + (\sum_j \beta_j - 1)Y_i - Z_i]\gamma_i$$

$$- Z_i x^i\beta = 0, \quad i = 1, 2, \ldots, q. \tag{11}$$

By solving equations (10) for $x^i\beta+\gamma_i$ and substituting the result into equations (8), there results, after some algebra,

$$x_{ij} = \frac{S_{ij}}{K_j + \beta_j \left(\frac{K_o - Z_i/\gamma_i}{1 - \sum_r \beta_r} \right)} - \gamma_i \tag{12}$$

$$i = 1, 2, \ldots, q;$$
$$j = 1, 2, \ldots, p.$$

Equations (9), (11), and (12) are the basis for an algorithm for the solution of the likelihood equation. The algorithm has the form

$$\tilde{Y}_i = \max(Y_i, 1)$$

$$\tilde{W}(0) = \operatorname{diag}(Y_1, \ldots, Y_q)$$

$$\tilde{\gamma}(0) = 1/K_o \, Z$$

$$\tilde{X}_{ij}(0) = \frac{S_{ij}}{K_j} - \tilde{\gamma}_i(0)$$

$$\hat{\beta}(k) = [\tilde{X}(k-1)\tilde{W}(k-1)^{-1}\tilde{X}(k-1)]^{-1} \tilde{X}^T(k-1)\tilde{W}(k-1)^{-1}[Y-\tilde{\gamma}(k-1)];$$

$$\tilde{\gamma}_i(k) = \Big[Y_i(1-\sum_j \hat{\beta}_j(k)) + Z_i - (K_o+1-\sum_j \hat{\beta}_j(k))\tilde{x}^i(k-1)\hat{\beta}(k)$$

$$+ \Big\{\Big(Y_i(1-\sum_j \hat{\beta}_j(k)) + Z_i - (K_o+1-\sum_j \hat{\beta}_j(k))\, \tilde{x}^i(k-1)\hat{\beta}(k)\Big)^2$$

$$+ 4(K_o+1-\sum_j \hat{\beta}_j(k))\, Z_i \tilde{x}^i(k-1)\hat{\beta}(k)\Big\}^{1/2}\Big]/2(K_o+1-\sum_j \beta_j(k)),$$

$$i = 1, \ldots, q;$$

$$\tilde{x}_{ij}(k) = \frac{S_{ij}}{K_j + \hat{\beta}_j(k)\left(\dfrac{K_o - Z_i/\tilde{\gamma}_i(k)}{1-\sum_r \hat{\beta}_r(k)}\right)} - \tilde{\gamma}_i(k),$$

$$i = 1, \ldots, q; \; j = 1, \ldots, p;$$

and

$$\tilde{W}(k) = \operatorname{diag}[\tilde{x}^1(k)\hat{\beta}(k)+\tilde{\gamma}_1(k), \ldots, \tilde{x}^q(k)\hat{\beta}(k)+\tilde{\gamma}_q(k)],$$

for $k = 1, 2, \ldots$. Then the maximum likelihood estimates are

$$\hat{\gamma} = \lim_{k \to \infty} \tilde{\gamma}(k)$$

$$\hat{X} = \lim_{k \to \infty} \tilde{X}(k)$$

and

$$\hat{\beta} = \lim_{k \to \infty} \hat{\beta}(k).$$

The asymptotic covariance matrix of the estimates is

$$\mathrm{COV}\begin{pmatrix}\beta\\ \gamma\\ x_1\\ \vdots\\ x_p\end{pmatrix} = V = \begin{pmatrix} S(\beta,\beta) & S(\beta,\gamma) & S(\beta,X)\\ S(\gamma,\beta) & S(\gamma,\gamma) & S(\gamma,X)\\ S(X,\beta) & S(X,\gamma) & S(X,X)\end{pmatrix}^{-1}$$

$$= \begin{pmatrix} V_{\beta\beta} & V_{\beta\gamma} & V_{\beta X}\\ V_{\gamma\beta} & V_{\gamma\gamma} & V_{\gamma X}\\ V_{X\beta} & V_{X\gamma} & V_{XX}\end{pmatrix}.$$

As before, the primary interest is in $V_{\beta\beta}$, which is again obtained by partitioning:

$$V_{\beta\beta} = \left\{S(\beta,\beta) - (S(\beta,\gamma)S(\beta,X)) \begin{pmatrix} S(\gamma,\gamma) & S(\gamma,X)\\ S(X,\gamma) & S(X,X)\end{pmatrix}^{-1} \begin{pmatrix} S(\gamma,\beta)\\ S(X,\beta)\end{pmatrix}\right\}^{-1}$$

Define

$$\lambda_{ij} = E[S_{ij}] = K_j(x_{ij}+\gamma_i)$$

$$\Lambda_j = K_j^{\,2}\ \mathrm{diag}(\lambda_{1j}, \lambda_{2j}, \ldots, \lambda_{qj})$$

$$\Lambda = \begin{pmatrix} \Lambda_1 & & & 0\\ & \Lambda_2 & & \\ & & \ddots & \\ 0 & & & \Lambda_p\end{pmatrix}$$

and

$$B = I_q \otimes \beta^T = (\beta_1 I_q, \beta_2 I_q, \ldots, \beta_p I_q).$$

Then

$$S(\beta,\beta) = X^T M^{-1} X$$

$$S(\beta,\gamma) = X^T M^{-1}$$

$$S(\beta,X) = X^T M^{-1} B$$

$$S(\gamma,\gamma) = M^{-1} + K_0 \Gamma^{-1} + \sum_j \Lambda_j^{-1}$$

$$S(\gamma,X) = M^{-1}B + (\Lambda_1^{-1}, \Lambda_2^{-1}, \ldots, \Lambda_p^{-1})$$

$$S(X,X) = \Lambda^{-1} + B^T M^{-1} B$$

so that

$$V_{BB} = \left\{ X^T \left[M^{-1} - (M^{-1}M^{-1}B) \begin{pmatrix} S(\gamma,\gamma) & S(\gamma,X) \\ S(X,\gamma) & S(X,X) \end{pmatrix}^{-1} \begin{pmatrix} M^{-1} \\ B^T M^{-1} \end{pmatrix} \right] X \right\}^{-1}.$$

Define

$$\begin{pmatrix} S(\gamma,\gamma) & S(\gamma,X) \\ S(X,\gamma) & S(X,X) \end{pmatrix}^{-1} = \begin{pmatrix} G_1 & G_2 \\ G_2^T & G_3 \end{pmatrix}.$$

Then

$$G_1 = [S(\gamma,\gamma) - S(\gamma,X)\, S(X,X)^{-1}\, S(X,\gamma)]^{-1}$$

$$G_2 = G_1\, S(\gamma,X)\, S(X,X)^{-1}$$

$$G_3 = S(X,X)^{-1} - S(X,X)^{-1}\, S(X,\gamma)\, G_1\, S(\gamma,X)\, S(X,X)^{-1}.$$

It can be shown that

$$S(X,X)^{-1} = \Lambda - \Lambda B^T M^{-1} (I^q + B\Lambda B^T M^{-1})^{-1}\, B\Lambda$$

and that G_1 is a diagonal matrix. Carrying out the considerable algebra involved in the indicated matrix operation leads to

$$V_{\beta\beta} = X^T W_*^{-1} X$$

where

$$W_* = M + \sum_j \left(\frac{\beta_j}{K_j}\right)^2 \Lambda_j + (1/K_o)\left(\sum_j \beta_j - 1\right)^2 \Gamma.$$

The maximum likelihood estimate of the asymptotic covariance $V_{\beta\beta}$ is obtained by substitution of "^" estimates for parameters in the definition of $V_{\beta\beta}$ in a manner similar to that for the known standards case.

The above result can also be written as

$$W_* = \text{diag}(W_1^*, W_2^*, \ldots, W_q^*)$$

where

$$W_i^* = \mu_i + \sum_j \left(\frac{\beta_j}{K_j}\right) \lambda_{ij} + (1/K_o)\left(\sum_j \beta_j - 1\right)^2 \gamma_i.$$

If this problem were to be treated as a weighted least squares problem using net sample and standard spectra, the appropriate weight for Y_i would be

$$\mathrm{Var}\left\{Y_i - \frac{Z_i}{K_o} - \sum_j \beta_j \left(\frac{S_{ij}}{K_j} - \frac{Z_i}{K_o}\right)\right\} = W_i^*, \quad i = 1, 2, \ldots, q.$$

Thus, the asymptotic covariance of the maximum likelihood estimates is the same as the covariance of the estimates obtained from an appropriate weighted least squares analysis.

AN EXAMPLE WITH THE STANDARD SPECTRAL SHAPES ASSUMED UNKNOWN

Iodine-129 is present in the environment from natural and man-made sources. The major man-produced sources are nuclear weapons tests and irradiated nuclear fuel processing plants. Because of its long half-life (1.6×10^7 years), ^{129}I is an excellent tracer for environmental processes involving other iodine isotopes. Environmental concentrations of ^{129}I and its ratio to natural iodine (^{127}I) can be studied at very low levels by neutron activation analysis of chemically separated iodine fractions. Neutron activation of ^{129}I and ^{127}I produces ^{130}I and ^{126}I. These latter two isotopes decay to radioxenon through a (β,γ) reaction. Thus ^{129}I concentration and the ratio $^{129}I/^{127}I$ can be estimated by measuring ^{126}I and ^{130}I gamma ray emission in an appropriate geometry (Brauer and Tenny, 1975).

The usual procedure for the radiochemical analysis of a neutron activated mixture of ^{126}I and ^{130}I is to fit the mixture energy spectrum to a linear combination of two standard spectra by weighted least squares. A complication is the difficulty in obtaining pure ^{126}I and ^{130}I standard spectra. Since ^{126}I has a 13-day half-life and ^{130}I has a 0.52-day half-life, an alternative to pure standards is two energy spectral shapes from radiochemical analysis at distinct points in time of a single solution of the two iodine isotopes. Weighted least squares, or in our case maximum likelihood, gives estimates of the amounts of the two temporally distinct mixtures in the sample spectrum. Knowledge of the relative disintegration rates of ^{126}I and ^{130}I in the standard mixture at a specific point in time (for example at reactor discharge of the iodine sample), the counting instrument geometries for the two

isotopes, and the isotopes' half-lives allows the writing of a one to one linear correspondence between the two mixture standards and the two pure standards. Hence, the mixture solution can be transformed into a solution for ^{126}I and ^{130}I.

What follows is a simple example of the above iodine radiochemical analysis for a very low level sample making use of two low level mixture standards.

A neutron activated iodine sample was counted for 185 minutes in a four-segment Na(I) five-inch well crystal using beta-gamma-gamma coincident circuitry. The resulting gross sample count versus energy spectrum is illustrated in Figure 1. A 300-minute background spectrum for this well crystal geometry is also illustrated in Figure 1. The background counts have been multiplied by (185/300) for direct comparison with the gross sample counts. To resolve the sample into ^{126}I and ^{130}I components, a mixture sample of ^{126}I and ^{130}I resulting from the neutron activation of a standard sample of known $^{129}I/^{127}I$ ratio was counted twice in the five-inch well crystal geometry at 1.249 and 3.060 days after reactor discharge. The counting intervals were 5 and 10 minutes, respectively. The two mixture spectra on a counts/minute basis are illustrated in Figure 2. In Figure 2 the two standard mixtures are almost identical in the first 25 energy bins and of similar shape but differing by a factor of 10 in energy bins 35 to 55. The low energy half of the spectrum then is primarily due to the 13-day longer lived isotope ^{126}I while the range from 35 to 55 energy bins is due to the 0.52-day isotope ^{130}I. In the standard mixture the $^{130}I/^{126}I$ disintegration rate ratio at reactor discharge, calculated from reactor physics equations, was 2.17. The well crystal geometries for the two isotopes were 0.106 and 0.493. Thus the counting rate ratio at reactor discharge was 0.467.

The sample and background spectra illustrated in Figure 1 and the two standard mixtures illustrated in Figure 2 were analyzed by the maximum likelihood procedure in the section on derivation of maximum likelihood estimates with standard spectra unknown. The counting data in energy bins 2 through 66 inclusive were used in the analysis. The maximum likelihood estimate of the two mixture spectra in total counts is

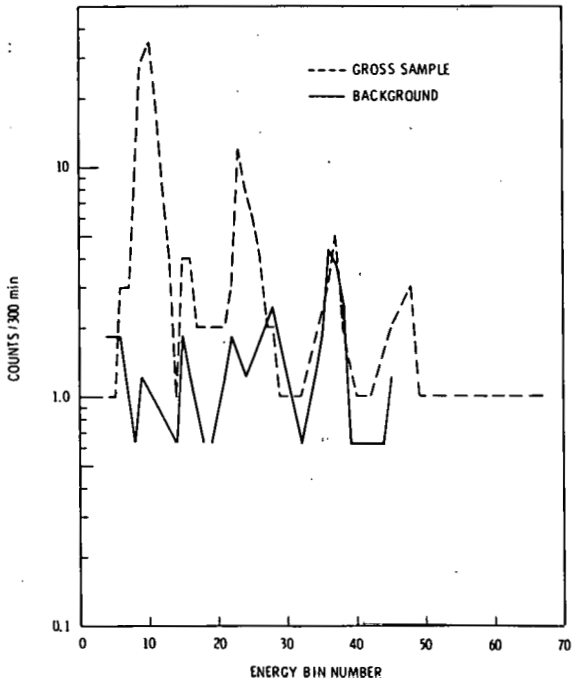


FIGURE 1. Gross Spectrum of Neutron Activated Iodine Sample and Background Spectrum for a Four-Segment Na(I) Five-Inch Well Crystal.

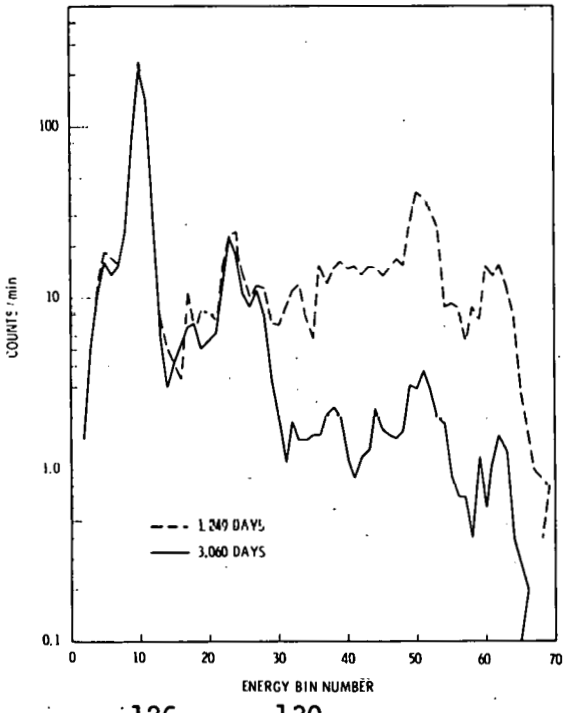


FIGURE 2. Spectra of ^{126}I and ^{130}I Counted 1.249 and 3.060 Days After Reactor Discharge of Neutron Activated Iodine Standard Sample.

$$\hat{\beta}_1 \sum_i \hat{x}_{i1} = -17.7 \qquad (13)$$

$$\hat{\beta}_2 \sum_i \hat{x}_{i2} = 164.$$

This estimate is on the boundary of the feasible space where $\hat{x}_{i1} \geq 0$, $\hat{x}_{i2} \geq 0$ and the two pure isotopic components of the mixture spectra $^{126}\hat{I}$ and $^{130}\hat{I}$ are non-negative. The pure isotopic estimates are related to equation (13) by the equation

$$^{n}\hat{I} = A(n,1)(\hat{\beta}_1 \sum_i \hat{x}_{i1}) + A(n,2)(\hat{\beta}_2 \sum_i \hat{x}_{i2}) \qquad (14)$$

with n = 126 and 130. Here,

$$A(n,j) = D_n e^{-\lambda_n T_j} / (D_n e^{-\lambda_n T_j} + D_m e^{-\lambda_m T_j})$$

where $D_n \neq D_m$ is the disintegration rate of ^{n}I at reactor discharge and T_j is the counting time of the j^{th} mixture standard (relative to reactor discharge time). From equations (13) and (14) the pure isotopic maximum likelihood estimates are

$$^{126}\hat{I} = 144.7 \text{ c/m}$$

$$^{130}\hat{I} = 0 \text{ c/m}.$$

The algorithm based on equations (9), (11) and (12) does not converge because the likelihood surface is not flat at the boundary point where the maximum occurs. The solution is obtained by exhaustive examination of a small region including the boundary area where the algorithm search exits the feasible parameter space. Since the likelihood surface is not flat at the maximum point, the asymptotic covariances, based on surface curvature, are not valid.

To get some idea of the precision of the solution, a section of the asymptotic 95% joint confidence for all estimated parameters is graphed in Figure 3. The section is defined by

$$\gamma = \hat{\gamma} \quad \text{and} \quad \mathbf{X} = \hat{\mathbf{X}}.$$

From Figure 3 the $^{126}\hat{I}$ estimate is clearly significant with an approximate 95% precision of 20^{+}%. The approximate 95% upper bound on ^{130}I is 1.6 c/m.

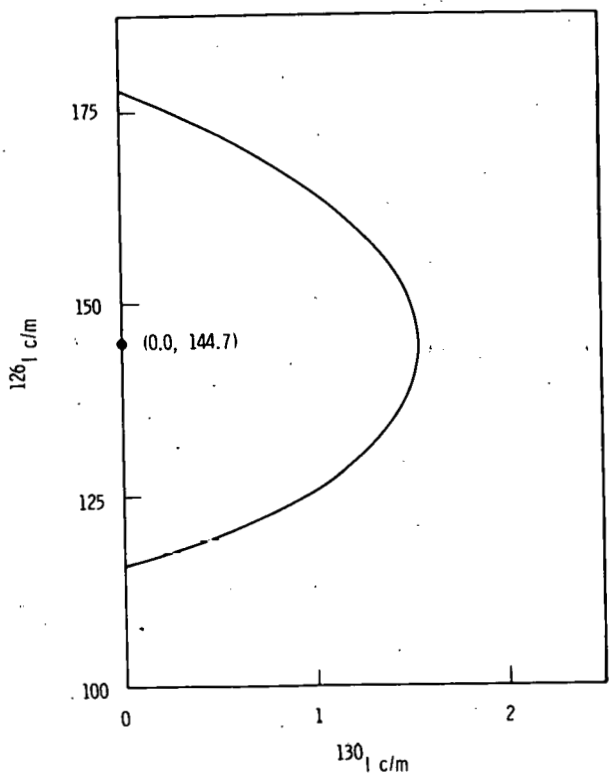


FIGURE 3. A Section Through Asymptotic Theory 95% Joint Confidence for Unknown Parameters Showing Contour for Iodine Isotopes.

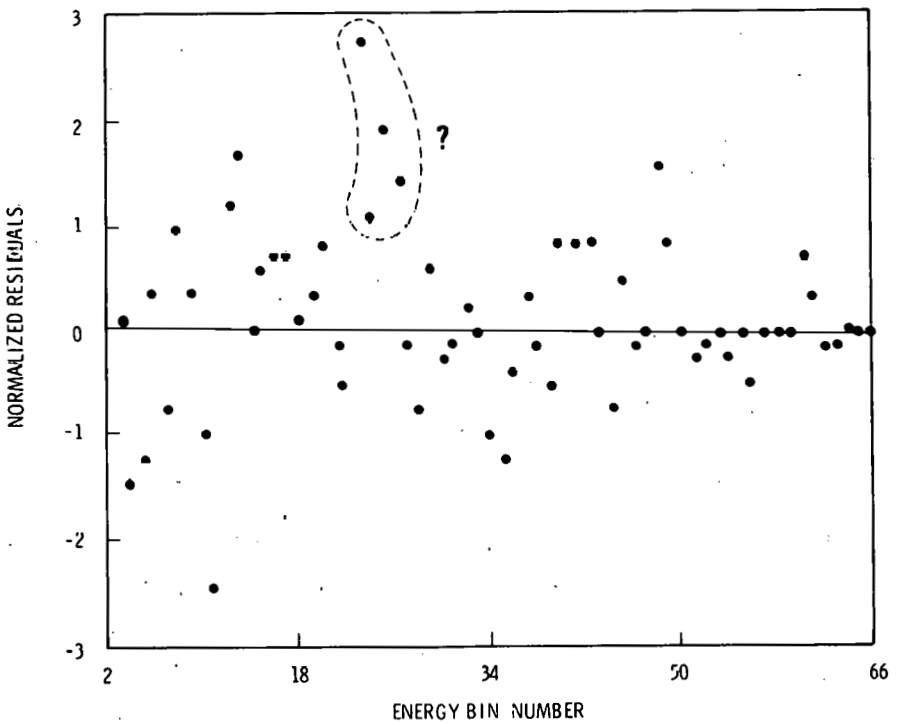


FIGURE 4. Normalized Residuals for Maximum Likelihood Estimate of Two Iodine Isotopes.

To assay the goodness of the maximum likelihood estimated model for gross sample counts, normalized residuals r_i can be calculated where

$$r_i = \sqrt{4Y_i+2} - \sqrt{4\mu_i+1} + \text{adjustment}_i .$$

This residual formula is suggested by Tukey (1976) for comparing Poisson counts Y_i to fitted counts μ_i. The addition of an adjustment term (Tukey, 1976) gives more appropriate residuals for low count situations. If no structure is left in the residuals (i.e., the model fits the data to within experimental error), they should vary about zero with standard deviation close to unity. While no exact distribution theory exists for determining when a residual is too big, a deviation of more than three or four from zero or a string of same sign residuals should arouse a question concerning agreement of data and model. These residuals do not necessarily behave like residuals from least squares fits to normally distributed data. Significance tests based on such theory are only approximate.

Adjusted normalized residuals are plotted versus energy bin number in Figure 4. There is no strong suggestion of model inadequacy. A second order effect is the four positive residuals (indicated by a question mark in Figure 4), all greater than 1.0 in bins 22 through 25. This is the region of a minor peak in each of the mixture standard spectra. Thus, there may be some systematic difference between the shape of peaks in standards and sample.

REFERENCES

Brauer, F. P., and Tenny, H., "^{129}I Analysis Method," BNWL-SA-5287, Battelle Pacific Northwest Laboratories, April 4, 1975.

Covel, D. F., Nucl. Instr. and Meth., 22 (1963), 101.

Horrocks, D. L., and Studier, M. H., "Determination of Radioactive Nobel Gases with a Liquid Scintillation," Anal. Chem., 36 (1964), 2077-79.

Jorgenson, D. W., "Multiple Regression Analysis of a Poisson Process," J. Amer. Statist. Assoc., 56 (1961), 235-245.

Nicholson, W. L., Schlosser, J. E., and Brauer, F. P., "Quantitative Analysis of Sets of Multicomponent Time-dependent Spectra from the Decay of Radionuclides," Nucl. Instr. and Meth., 25 (1963), 45.

Pasternack, B. S., "Linear Estimation and the Analysis of Gamma-ray Pulse-height Spectra," Technometrics, 4 (1962), 565-571.

Pasternack, B. W., and Liuzzi, A., "Patterns in Residuals: A Test for Regression Model Adequacy in Radionuclide Assay," Technometrics, 7 (1965), 603-621.

Rao, C. R., Linear Statistical Inference and Its Application, John Wiley and Sons, Inc., New York, 1965.

Salmon, L., "Computer Analysis of Gamma Ray Spectra from Mixtures of Known Nuclides by the Method of Least Squares," Proceedings of a Symposium on Application of Computers to Nuclear and Radiochemistry, Gatlinburg, Tennessee, October 17-19, 1962, NAS-NS 3107, (1963), 165-183.

Schonfeld, K., Kibbey, A. H., and Davis, W., Jr., "Determination of Nuclide Concentrations in Solutions Containing Low Levels of Radioactivity by Least Squares Resolution of the Gamma-ray Spectra," Nucl. Instr. and Meth., 45 (1966), 1-21.

Slavinskas, D. D., Kennett, T. J., Prestwich, W. V., and Keech, G. L., "Analysis of Two-dimensional Coincidence Spectra," Nucl. Instr. and Meth., 41, 1966, 1-12.

Shumway, R. H., Radionuclide Analysis of Gamma-ray Spectra by Stepwise Multiple Regression, Public Health Service Publication No. 999-R-5, 1963.

Tukey, J. W., Exploratory Data Analysis, First Edition, Addison-Wesley Publishing Co., Reading, MA, to appear Fall 1976.

Turkevich, A. L., Franzgrote, E. J., and Patterson, J. H., "Chemical Analysis of the Moon at the Surveyor V Landing Site," Science, 158 (1967), 635-637.

von Neumann, J., "Distribution of the Ratio of the Mean Square Successive Difference to the Variance," Annals of Mathematical Statistics, 12 (1941), 367-395.

Wilks, S. S., Mathematical Statistics, John Wiley and Sons, Inc., New York, 1962.

Yamamoto, S., "Least Squares Analysis of Time-dependent Gamma-ray Pulse-Height Spectra," Nucl. Instr. and Meth., 52 (1967), 245-249.

A QUALITY CONTROL PROGRAM FOR 100% INSPECTION USING NON-DESTRUCTIVE MEASUREMENTS

R. L. Hooper
Battelle Pacific Northwest Laboratories
Richland, Washington

A quality control program that allows the producer some flexibility in allocating program resources defines consumer risk as the maximum probability of accepting a defect of limiting size. Control chart sample sizes and product acceptance criteria are computed as a function of consumer risk. Results appear in a convenient nomograph where the vendor can optimize the choice of control chart sample size and accept criteria relative to a particular measurement system while maintaining the appropriate level of consumer risk.

A quality control program is developed for the acceptance of individual items where 100% inspection is accomplished using a non-destructive measurement system. The program is a function of the measurement system's response to defects of a known size (standards), measurement precision and a continuing plan for demonstrating that the measurement system is in control. Naturally, the major aim of the quality control program is to protect the quality of the product delivered to the consumer. However, one of the goals is to provide the producer with some flexibility in terms of how he dedicates his resources in support of the program.

The measurement system is initially calibrated using certified standards. The levels of the standard defects should span the limiting levels which are specified for the characteristics in question. The calibration exercise is designed to accomplish the following:

(1) Define the tester response to defects of known size.
(2) Provide information necessary to develop an estimate of tester precision.
(3) Define the stability of tester response across a time period similar to the period that the tester is employed during acceptance testing.

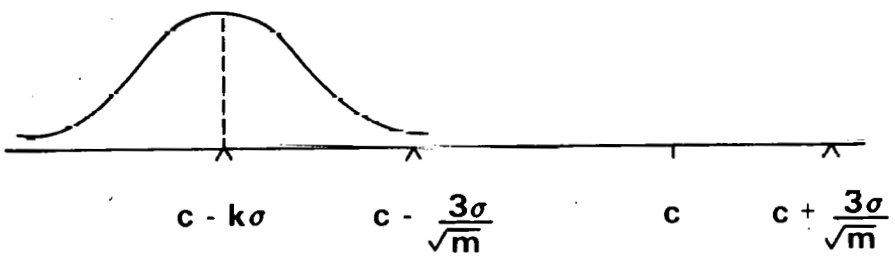
Periodically, at intervals established on the basis of stability and tester usage, the standard defects will be submitted for measurement and the results maintained in a control chart format. The data generated during the calibration exercise can also be used to examine the viability of the assumptions necessary for the development that follows. The assumptions are:

(1) The non-destructive tester response is linear across the range of defects observed in application.
(2) The distribution of tester response for a given defect is normal.
(3) The variance is the same for all defects.

The consumer risk is defined as the maximum probability of a defect of limiting size. It is computed as the max $\{P_1 P_2\}$ where:

P_1 = probability of accepting the tester as being correctly calibrated.
P_2 = probability of accepting a defect of limiting size

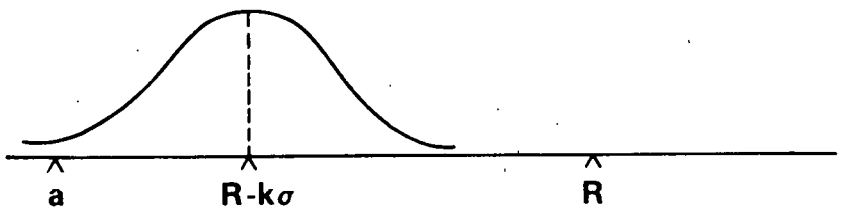
Under the assumption that the distribution of tester responses to a defect of given size is normal and the variance of tester responses is the same for all defects of interest, then the probabilities, P_1 and P_2, are computed directly as shown in Figures 1 and 2.



$$P_1 = \int_{c-\frac{3\sigma}{\sqrt{m}}}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma/\sqrt{m}}\, e^{-\frac{1}{2\sigma^2/m}(x-c+k\sigma)^2}\, dx = \int_{c-\frac{3\sigma}{\sqrt{m}}}^{\infty} n(x \mid c-k\sigma,\ \sigma/\sqrt{m})\, dx$$

c - AVERAGE TESTER RESPONSE TO A STANDARD DEFECT
m - CONTROL CHART SAMPLE SIZE

FIGURE 1. Three-sigma Control chart



$$P_2 = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x - R+k\sigma)^2} dx = \int_{-\infty}^{a} n(x|R-k\sigma, \sigma)\, dx$$

R - NOMINAL TESTER RESPONSE TO A DEFECT OF LIMITING SIZE.
a - ACCEPTANCE CRITERIA APPLIED IN PRODUCTION TESTING

FIGURE 2. Tester Response to Defect of Limiting Size

Now the maximum probability of accepting a defect of limiting size = $\max_k \{P_1 P_2\}$. The product is maximum when the $\partial P_1 P_2/\partial k$ is zero

$$\frac{\partial P_1 P_2}{\partial k} = P_1 \frac{\partial P_2}{\partial k} + P_2 \frac{\partial P_1}{\partial k} = 0 \qquad (1)$$

From above, following simplification, we have

$$P_1 = \int_{c - \frac{3\sigma}{\sqrt{m}}}^{\infty} n(x|c-k\sigma, \sigma/\sqrt{m})\, dx = 1 - N[\sqrt{m}\, k - 3]$$

$$\frac{\partial P_1}{\partial k} = \frac{\partial}{\partial k} \int_{c - \frac{3\sigma}{\sqrt{m}}}^{\infty} n(x|c-k\sigma, \sigma/\sqrt{m})\, dx = -\sqrt{m}\, n[\sqrt{m}\, k - 3|0,1]$$

$$P_2 = \int_{-\infty}^{a} n(x|R-k\sigma,\ \sigma)dx = N(k - \frac{R-a}{\sigma})$$

$$\frac{\partial P_2}{\partial k} = \frac{\partial}{\partial k}\int_{-\infty}^{a} n(x|R-k\sigma,\ \sigma)\ dx = n[k - \frac{R-a}{\sigma}\ \Big|\ 0,1]$$

Substitution yields

$$[1 - N(\sqrt{m}\ k-3)]\ n\left[k - \frac{R-a}{\sigma}\ |0,1\right] - \sqrt{m}\left[N\left(k - \frac{R-a}{\sigma}\right)\right] n[\sqrt{m}\ k - 3|0,1] = 0$$

and

$$\frac{\sqrt{m}[n(\sqrt{m}\ k - 3\ |\ 0,1)]}{1 - N(\sqrt{m}\ k-3)} = \frac{n\left[k - \frac{R-a}{\sigma}\ 0,1\right]}{N(k - \frac{R-a}{\sigma})} \qquad (2)$$

Equation (2) is solved for a range of values of $\frac{R-a}{\sigma}$ (the standardized distance that the acceptance value (a) is removed from the nominal tester response (R) to a defect of limiting size) and selected values of m. The results are shown in nomograph form in Figure 3. Consumer risk is described on the ordinate axis.

The producer, with a given measurement system, can optimize his choice of a and m and still maintain the required consumer risk. The producer should select the largest value of "a" (minimum producer risk) consistent with the financial penalty associated with an increasing control chart sample size, m. These results plus a number of examples appear as an RDT Standard (RDT F 3-13).

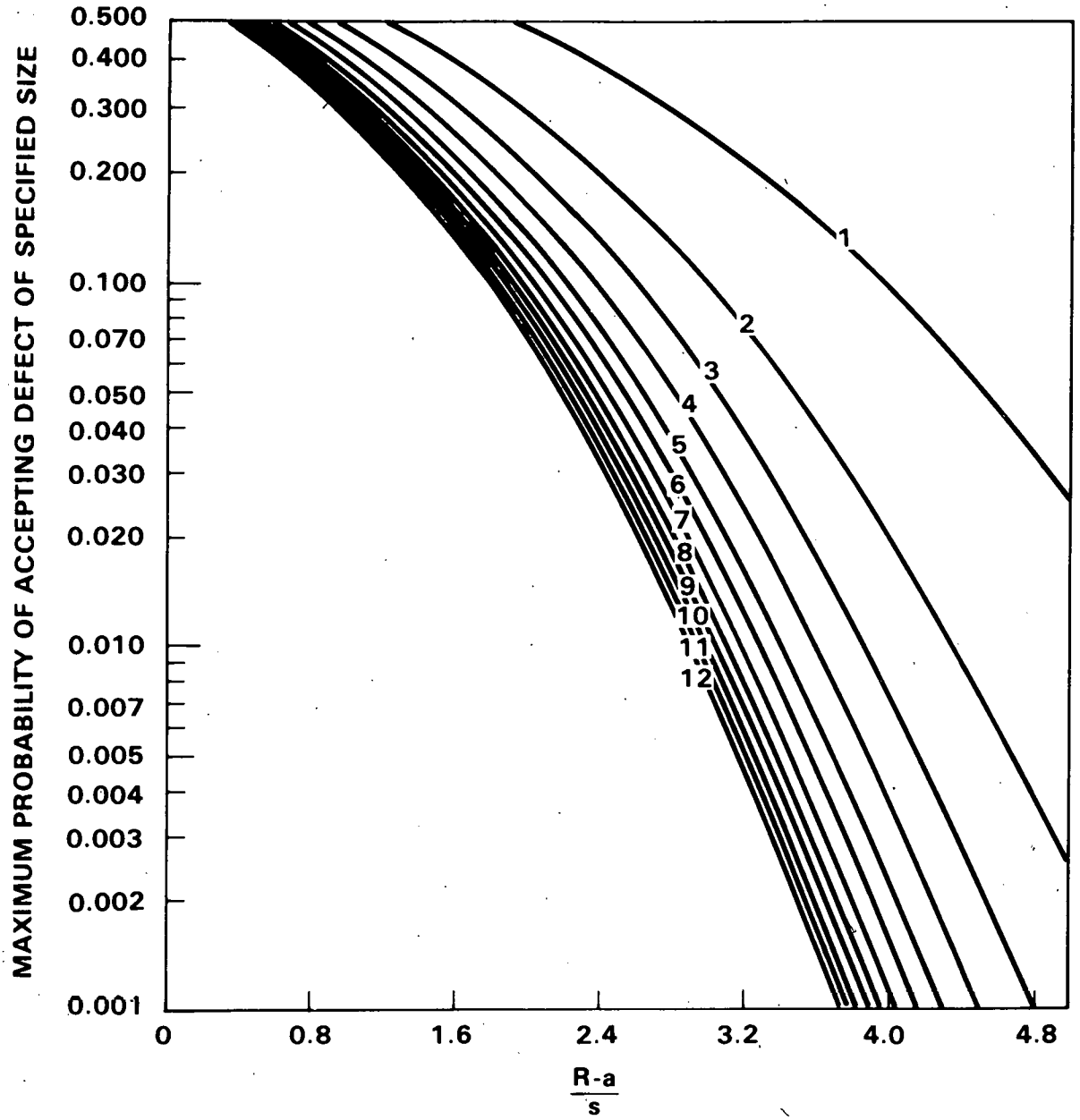


FIGURE 3. Control Chart Sample Size and Product Acceptance Criteria

DISTRIBUTION OF SKEWNESS AND KURTOSIS STATISTICS

K. O. Bowman
Union Carbide Corporation Nuclear Division
Oak Ridge, Tennessee

and

L. R. Shenton
University of Georgia
Athens, Georgia

Tests which allow for both skewness and kurtosis ("omnibus" tests) have been studied relative to non-normal sampling. These investigations have led to acceptance contours at levels of 90, 95 and 99% for sets of sample sizes over populations from both Pearson system and normal mixtures (two-components, equal variances). Ongoing studies relate to a closer examination of the joint distribution of $\sqrt{b_1}$ and b_2 with the possibility of producing contours with less area and fixed probability content.

INTRODUCTION

EARLY WORK

The skewness (defined as the standardized third central sample moment $\sqrt{b_1} = m_3/m_2^{3/2}$) and kurtosis ($b_2 = m_4/m_2^2$) have a long history as statistical tests for departures from normality, spanning the greater part of the last 100 years. In addition, the population values $\sqrt{\beta_1}$, β_2 form the basis of the orderly used Pearson system of distributions. It is clear that $\sqrt{b_1}$, depending on cubes of deviation, is a measure of skewness; it is customary to interpret b_2 as a measure of 'peakedness' or lack of it ('flatness'). If the Pearson ($\sqrt{\beta_1}$, β_2) diagram of curve types (I, III, IV, V being the main members) is recalled, large values of β_2 are associated with Type IV curves (including Student's t for example) which have long tails or a slow to very slow approach to zero frequency; equivalently, this implies a relatively large chance of a large deviation from the mean. Similarly, small values of β_2 are associated with Type I (or Beta) distribution having

finite range (including, for example, the uniform distribution). The pivotal distribution is the normal for which $\sqrt{\beta_1} = 0$, $\beta_2 = 3$.

Of the classical test statistics, Student's t, F-ratio, correlation coefficient (and perhaps mean and variance), the skewness and kurtosis statistics are the only ones whose distributions in a normal sampling are still not known exactly. Could the reason for this be that the other quantities are dominated by quadratic forms (including degenerate cases), whereas $\sqrt{b_1}$, b_2 involve higher forms?

SERIES DEVELOPMENTS

An important breakthrough came from E. S. Pearson (1930), who used the work of Fisher (1928) and Wishart (1930) on k-statistics to develop a Taylor series expansion in terms of the k-statistic discrepancies $k_i - \kappa_i$, i=2,3,4. To damp-out higher order terms, Pearson used samples of $n \geq 50$ for $\sqrt{b_1}$, $n \geq 100$ for b_2 to assess the lower and upper 1% and 5% of the distributions in normal sampling. Thirty or so years later (Pearson, 1965), he gave a set of 'accepted' percentage points; for a sample of 50 there is no change in the third d.p. entries for $\sqrt{b_1}$ at the 1%, 5% levels; for b_2 and n=100 there is no change in the second d.p. entries; in all, quite a remarkable achievement.

EXACT MOMENTS

The next step forward came from Fisher (1930) who showed that in normal sampling the standardized moments $m_r/m_2^{r/2}$, $r \geq 3$, are distributed independently of the second moment m_2 (Fisher used k-statistic notation). Thus, for example, $Eb_1 = Em_3^2/Em_2^3$ follows from the independence of m_3^2/m_2^3 and m_2^3. In this manner, the exact moments of $\sqrt{b_1}$ and b_2 can be found. In fact, Fisher derived the first six cumulants of $\sqrt{b_1}$ and the first four cumulants of b_2. Pepper (1932) derived the eighth moment of $\sqrt{b_1}$, Hsu and Lawley (1939) the fifth and sixth moments of b_2. Later, Geary and Worlledge (1947) gave the seven noncentral moment of b_2; actually they give $E(m_4/m_2^2)^7$. Some of the coefficients are quite large, that of n^3 being a 13-digit integer multiplied by 25515 (the whole expression has, so far, had scant usage).

APPROXIMATING DISTRIBUTIONS

Knowing exact moments, it was a natural development to search for approximating distributions, reaching out towards percentage points of the distributions. Four-moment fits were studied by Pearson (1963) and at this time he could use the Pearson system, Gram-Charlier series based on the normal, and the Johnson S_U translation system (Johnson and Kotz, 1970). For $n \geq 30$, the Student-t density gave an acceptable approximation for $\sqrt{b_1}$, the criterion being the closeness of agreement between the standardized sixth and eighth moments for the model and the true values. Johnson's S_U, although being troublesome to fit, seemed to be equally acceptable to Pearson (1963, p. 106). Recently, D'Agostino (1970) has shown that Johnson's S_U for $n \geq 8$ gives a very acceptable approximation in this case.

For the kurtosis, the problem is more difficult because the statistic is one-sided and very skew in general; thus for n=25, $\sqrt{\beta_1}(b_2) = 1.75$, $\beta_2(b_2) = 9.90$ in comparison to $\sqrt{\beta_1}(\sqrt{b_1}) = 0$, $\beta_2(\sqrt{b_1}) = 3.58$. Briefly, Pearson Types VI and IV or Johnson's S_U can be regarded as acceptable approximations to the distributions of b_2 for $n \geq 40$ according to unpublished work of C. T. Hsu (quoted by Pearson). Hsu pointed out that for $n \geq 30$, the (β_1,β_2) points for b_2 were close to the Type V line (Pearson, 1963, p. 106). Noting this, Anscombe and Glynn (1975) in a Technical Report made available at the Atlanta ASA meeting (August 1975), suggest a linear function of a reciprocal of a χ^2-variate as an approximation to b_2 for $n > 30$ or so; they do not make any comparisons with Johnson's S_U approximation.

OBJECTIVES

For some time, we have been tackling the problem of finding moments of sample moments which result in Taylor-series developments in powers of the inverse sample size. The first objective was to derive a technique to generate the terms, and having dealt with this by a computer approach, the next problem relates to the use of the locked-in information. This second phase is still a problem area.

An obvious application concerns the distributional properties of $\sqrt{b_1}$, b_2 in general sampling, and pushing this to extremes would involve a knowledge of the joint distribution. A related problem which has been considered is to try to decide from sample data whether the sampling is from a specified Pearson distribution (or other general population). This would, one imagines, require information on the sampling distribution of the Pearson criteria $\kappa_1 = 2b_2 - 3b_1 - 6$ and

$$\kappa_2 = b_1(b_2 + 3)^2[4(4b_2 - 3b_1)(2b_2 - 3b_1 - 6)]^{-1}$$

where, for example, for Type I, κ_1 and $\kappa_2 < 0$.

This explains our interest in the marginal distributions of $\sqrt{b_1}$, b_2 and joint moments in general sampling. In the sequel we refer to various related topics, and if the account seems somewhat disjointed, it reflects a natural tendency, when confronted with a seemingly difficult problem. to try something simpler.

The problem relating to κ_1, κ_2 is still unresolved, but at least we have a better understanding of the early phase problems and also some doubts as to whether it is indeed resolvable.

NON-NORMAL SAMPLING

MOMENTS OF SAMPLE-MOMENTS

The independence property of the skewness (and kurtosis) and variance breaks down in sampling from Pearson (except the normal) and other distributions. Using Taylor series developments, the present authors along with D. Sheehan (1971) set up recursive schemes for the moments of sample moments (such as $\sqrt{m_2}$, Student's-t, etc.) suitable for programming for a computer. The flavor of the results is contained in the following cases:

(a) Univariate. If

$$A_s = E(m_1' - \mu_1')^s \ , \ a_s = E(X - \mu_1')^s \ ,$$

then

$$A_s^{(k)} = \sum_{r=1}^{s} \binom{s}{r} \{a_{r+1} A_{s-r}^{(k-r)} - a_r A_{s-r+1}^{(k-r)}\}$$

where $A_s^{(k)} = 0$ for $k \geq s$, $k < 0$; and it refers to the coefficient of n^{-k} in A_s.. In particular,

$$A_2 = \mu_2/n \quad , \qquad A_3 = \mu_3/n^2 \quad ,$$

$$A_4 = 3a_2^2/n^2 + (a_4 - 3a_2^2)/n^3 \quad .$$

(b) Bivariate.

$$A_{s+1,t}^{(k)} = \sum_{\lambda=0}^{s} \sum_{\mu=0}^{t} \binom{s}{\lambda}\binom{t}{\mu} a_{s+1-\lambda,t-\mu} A_{\lambda,\mu}^{(k+\lambda+\mu-s-t)} \qquad (2)$$

$$- \sum_{\lambda=0}^{s} \sum_{\mu=0}^{t} \delta_{\lambda,\mu} \binom{s}{\lambda}\binom{t}{\mu} a_{\lambda,\mu} A_{s+1-\lambda,t-\mu}^{(k-\lambda-\mu)},$$

$$A_{s,t+1}^{(k)} = \sum_{\lambda=0}^{s} \sum_{\mu=0}^{t} \binom{s}{\lambda}\binom{t}{\mu} a_{s-\lambda,t+1-\mu} A_{\lambda,\mu}^{(k+\lambda+\mu-s-t)}$$

$$- \sum_{\lambda=0}^{s} \sum_{\mu=0}^{t} \delta_{\lambda,\mu} \binom{s}{\lambda}\binom{t}{\mu} a_{\lambda,\mu} A_{s-\lambda,t+1-\mu}^{(k-\lambda-\mu)},$$

where $[(s+t+1)/2] \leq 1 \leq s+t$, $A_{s,t}^{(k)} = 0$ for $k < 0$. or $k \geq s+t$, $\delta_{\lambda,\mu} = 1$ unless $\lambda = \mu = 0$; $\delta_{\infty} = 0$, and $[x]$ indicates the largest integer $\leq x$. $\underline{A}$ and a have similar meanings as in (a); for example,

$$A_{r,s} = E(m_1' - \mu_1')^r (m_2' - \mu_2')^s \quad ,$$

$$a_{r,s} = E(X - \mu_1')^r (X^2 - \mu_2')^s$$

would be a possibility, or similar expression involving linear sums of non-central sample moments and the corresponding expressions for samples n=1.

Using this approach, we have set up, using a computer, the first eight moments of $\sqrt{b_1}$ and the first six moments of b_2 up to and including the term in n^{-8} in the sample size. Momments of $\sqrt{b_1}$ involve four-dimensional arrays (corresponding to the first three sample moments and the sample size); similarly those of b_2 involve five-dimensional arrays. As for the moments of the population sampled, the first 40 are needed for

the eight moments of $\sqrt{b_1}$, and the first 44 for the six moments of b_2; it is preferable to set up recursive schemes for these.

ILLUSTRATIONS

(a) Population Uniform Distribution ($\sqrt{\beta_1} = 0$, $\beta_2 = 1.8$)

$$\mu_2(\sqrt{b_1}) \sim \frac{2.0571}{n} + \frac{2.2629}{n^2} + \frac{1.8042}{n^3} + \ldots \frac{-5.2943^{(5)}}{n^8}$$

$$\mu_4(\sqrt{b_1}) \sim \frac{1.2696^{(1)}}{n^2} + \frac{4.7687^{(1)}}{n^3} + \ldots \frac{-4.6669^{(5)}}{n^8}$$

$$\mu_6(\sqrt{b_1}) \quad \frac{1.3058^{(2)}}{n^3} + \ldots + \frac{1.0269^{(6)}}{n^8}$$

$$\mu_8(\sqrt{b_1}) \sim \frac{1.8804^{(3)}}{n^4} + \ldots + \frac{1.8986^{(7)}}{n^8}$$

(Superscripts in parentheses refer to the power of 10 to be used as a multiplier.)

(b) Population Pearson Type I, $\sqrt{\beta_1} = 0.2$, $\beta_2 = 3.1$

$$E\sqrt{b_1} \sim 0.2 - \frac{1.4211}{n} + \frac{6.5992}{n^2} - \frac{3.9728^{(1)}}{n^3} + \frac{3.9971^{(2)}}{n^4} - \frac{5.9343^{(3)}}{n^5}$$

$$+ \frac{1.487^{(5)}}{n^6} - \frac{2.7672^{(6)}}{n^7} + \frac{8.0796^{(7)}}{n^8}$$

$$\text{Var } \sqrt{b_1} \sim \frac{6.7573}{n} - \frac{6.1810^{(1)}}{n^2} + \ldots - \frac{8.5309^{(9)}}{n^9}$$

$$\mu_8(\sqrt{b_1}) \sim \frac{2.1891^{(5)}}{n^4} + \ldots - \frac{1.6024^{(15)}}{n^8}$$

(c) Population Normal Mixture (2-components, equal variances)

$\sqrt{\beta_1} = 1.0 \quad \beta_2 = 4.0$

$$E\, b_2 \sim 4 - \frac{2.3142}{n} - \frac{9.7524^{(1)}}{n^2} + \ldots - \frac{4.0862^{(8)}}{n^8}$$

$$\mu_4(b_2) \sim \frac{8.5535^{(3)}}{n^2} + \frac{5.2878^{(5)}}{n^3} + \ldots - \frac{5.390^{(12)}}{n^8}$$

$$\mu_6(b_2) \sim \frac{2.2836^{(6)}}{n^3} + \ldots - \frac{2.5635^{(15)}}{n^8}$$

(d) Population Pearson Type I, $\sqrt{\beta_1} = 1.0$, $\beta_2 = 4.0$

$$E(b_2\sqrt{b_1}) \sim 4 - \frac{9.7966}{n} - \frac{4.9612^{(2)}}{n^2} + \ldots + \frac{5.7413^{(7)}}{n}$$

$$\mathrm{Var}(b_2\sqrt{b_1}) \sim \frac{5.6987^{(2)}}{n} + \frac{3.9918^{(3)}}{n^2} + \ldots + \frac{3.4572^{(10)}}{n^6}$$

$$\mu_3(b_2\sqrt{b_1}) \sim \frac{2.9913^{(5)}}{n^2} + \ldots - \frac{8.0078^{(11)}}{n^6}$$

$$\mu_4(b_2\sqrt{b_1}) \sim \frac{9.7426^{(5)}}{n^2} + \ldots - \frac{9.9575^{(13)}}{n^6}$$

SAFE SAMPLE SIZE

Our early work in using these expansions to approximate the distributions of, for example, $\sqrt{b_1}$ and b_2, relied on inflating the sample size to damp-out higher order terms. Thus (Bowman and Shenton, 1975) safe sample sizes are indicated for each moment, using the rather arbitrary rule that the critical sample size is one which adjusts the size of the highest order to the lowest order terms to be approximately one-tenth. For example, in sampling from Type I with $\sqrt{\beta_1} = 0.6$, $\beta_2 = 3.2$, the safe sample size for $E\sqrt{b_1}$ is n = 10,

$$E\sqrt{b_1} \sim 0.6 - 0.3082 + 0.0444 + 0.0227 - 0.0302 + 0.0175 + 0.0186 - 0.0606 + 0.0414 \ ;$$

Similarly, in sampling from Pearson Type I with $\sqrt{\beta_1} = 1.4$, $\beta_2 = 3.4$, the critical size is n = 100 for $\mu_5(b_2)$ and

$$\mu_5(b_2) \sim 3.4419 + 3.7942 + 2.5590 + 1.4111 + 0.7088 + 0.3417 \ .$$

Clearly, in both cases the sample sizes are only just adequate to damp-out the n^{-8} terms. Pearson (1930) used rather similar damping factors; for example, for normal samples he gave

$$\sigma(\sqrt{b_1}) \sim 0.3464 \ (1 - .0600 + .0024 - .0001)$$

when n = 50, and

$$\beta_2(b_2) \sim 3 + 5.4000 - 2.0196 + 0.4704$$

when n = 100. In the case of σ it looks as if a smaller sample size could have been used.

RATIONAL FRACTION AND OTHER APPROXIMATIONS

Asymptotic or slowly convergent series may be approximated by the ratio of polynomials in the variable (in our case, n the sample size), and there has been a resurgence of interest in the last decade in the subject, basically initiated by Padé (his thesis was published in 1892). Briefly, the domain of convergence of Padé approximants (which include Stieltjes continued fractions as a special case) is generally more extensive than is the case for series developments; for series in 1/n, this suggests the possibility that smaller values of n may be valid in Padé approximants. Genuinely divergent series (or what appear to be so from the pattern of the first few terms) seem to be quite common in statistics; at least that is our experience, but from a knowledge of a few terms (8, 15, or perhaps 30), one cannot claim they are authentic examples (Van Dyke, 1974, p. 436). An account of the recent prolific developments in the subject is to be found in Baker (1975) and Baker, Jr., and Gammel (1970).

We consider a few examples:

- $E\sqrt{b_1}$ from Pearson Type I with $\sqrt{\beta_1} = 0.2$, $\beta_2 = 3.1$

The first four approximants $p_i = P_i(n)/Q_i(n)$, i=1,2,3,4 where P_i, Q_i are polynomials in n of degrees i-1, i, respectively, are

n = 10: $P_1 = 0.1169$ $P_2 = 0.0968$ $P_3 = 0.1016$ $P_4 = 0.1018$

n = 20: $P_1 = 0.1476$ $P_2 = 0.1415$ $P_3 = 0.1420$ $P_4 = 0.1420$

Similarly, using rational fraction approximants for the n^{-1} term onwards and adjusting for the constant term, we have:

n = 10: $p_1^* = 0.1029$ $p_2^* = 0.1014$ $p_3^* = 0.1014$ $p_4^* = 0.1020$

n = 20: $p_1^* = 0.1423$ $p_2^* = 0.1419$ $p_3^* = 0.1419$ $p_4^* = 0.1420$

For n = 10 the series is useless; for n = 20 it gives 0.1439.

- Eb_2 from Pearson Type I with $\sqrt{\beta_1} = 1$, $\beta_2 = 2.4$

$$Eb_2 \sim 2.4 + \frac{7.9069}{n} + \frac{24.6491}{n^2} + \frac{43.1163}{n^3} - \frac{1144.60}{n^4}$$

$$- \frac{34287.9}{n^5} - \frac{7.8965^{(5)}}{n^6} - \frac{1.7720^{(7)}}{n^7} - \frac{4.0550^{(8)}}{n^8}$$

n = 10: $p_1 = 3.5488$ $p_2 = 3.5783$ $p_3 = 3.7228$ $p_4 = 3.8332$

Series - -3.59

n = 20: $p_1 = 2.8682$ $p_2 = 2.8765$ $p_3 = 2.9844$ $p_4 = 2.7771$

n = 50: $p_4 = 2.5680$ Series = 2.5680

It is well known that just as any finite number of terms of a Taylor series cannot in general bound the true value, similarly there are problems in interpreting Padé sums. However, the method does hold promise for making available acceptable (in some sense) approximations to moment series (when the sample size is relatively small) with seemingly divergent tendencies. We are studying other methods of summation (generalized Padé and Borel-Padé).

THE CORRELATION BETWEEN $\sqrt{b_1}$ AND b_2

ASYMPTOTIC CORRELATION

Early work goes back to Pearson (1902) who gave, for general sampling, the first-order asymptotics for $\mu_2(b_1)$, $\mu_2(b_2)$, and the correlation

$$R(b_1,b_2) = \frac{(2\beta_6-3\beta_1\beta_4-4\beta_2\beta_3+6\beta_1\beta_2^2+3\beta_1\beta_2-6\beta_3+12\beta_1^2+24\beta_1)}{[\mu_2(b_1)\mu_2(b_2)]^{1/2}} \qquad (3)$$

where

$$\beta_3 = \mu_3\mu_5/\mu_2^4 , \quad \beta_4 = \mu_6/\mu_2^3 , \quad \beta_5 = \mu_3\mu_7\mu_2^5 , \quad \beta_6 = \mu_8/\mu_2^4 .$$

Note, as far as first-order terms, this is the same as the correlation between $\sqrt{b_1}$ and b_2

MORE EXACT RESULTS

Further coefficients in higher powers of n^{-1} can be used in

$$\rho(\sqrt{b_1},b_2) = \frac{E(b_2\sqrt{b_1}) - Eb_2E\sqrt{b_1}}{\sigma(b_2)\ \sigma(\sqrt{b_1})}$$

by the method of moments of sample moments.

Actually, one of us used a computer to approximate ρ, the series for $E(b_2\sqrt{b_1})$ being taken as far as n^{-6}, those for $\sqrt{b_1}$ and b_2 to n^{-8}. A selection of results is given in Table 1. The interesting aspect is the high correlation for samples of 50 upwards, which increases with sample size, but not markedly. In addition, although the population sampled does have an effect, the correlation is still high and quite stable over sample size. This raises a question as to the role likely to be played by b_2 in a test of a hypothesis that a population sampled was a Pearson distribution with given $\sqrt{\beta_1}$, β_2. The property does also throw a little light on the joint distribution of $\sqrt{b_1}$, b_2 in non-normal sampling.

TABLE 1. Correlation Between $\sqrt{b_1}$ and b_2

$\sqrt{\beta_1}$	β_2		Sample Size 50	75	100	250	500	1000
0.2	1.2	(a)*	0.716	0.775	0.811	0.891	0.924	0.942
		(b)	0.696	0.748	0.779	0.847	0.874	0.889
0.6	2.6	(a)	0.756	0.761	0.764	0.768	0.770	0.771
		(b)	0.819	0.824	0.827	0.832	0.834	0.834
0.6	3.4	(a)	0.638	0.642	0.643	0.645	0.645	0.646
		(b)	0.834	0.832	0.836	0.845	0.847	0.849
0.6	4.0	(a)	0.696	0.704	0.707	0.711	0.711	0.711
		(b)	--	--	--	--	--	--
1.0	2.2	(a)	0.965	0.977	0.983	0.992	0.995	0.997
		(b)	0.965	0.975	0.980	0.989	0.992	0.993
1.0	2.6	(a)	0.952	0.963	0.968	0.976	0.979	0.980
		(b)	0.953	0.960	0.964	0.970	0.972	0.973
1.0	3.8	(a)	0.768	0.774	0.779	0.778	0.792	0.793
		(b)	0.911	0.913	0.914	0.915	0.916	0.916

*Populations sampled: (a) Normal Mixture, (b) Pearson Type I. Moments based on asymptotic series); Pearson Type I does not apply to the case (0.6, 4.0).

SIMULTANEOUS BEHAVIOR OF $\sqrt{b_1}$, b_2

OMNIBUS TESTS

If one accepts the notion that Johnson's S_U provides a good approximation to the distributions of $\sqrt{b_1}$ and b_2, then we are able to derive approximate equivalent normal deviates in each case. Now S_U is defined by the relation

$$X = \gamma + \delta \sinh^{-1}\left(\frac{x - \zeta}{\lambda}\right), \qquad (X \in N(0,1)) \tag{4}$$

where, for example, x refers to $\sqrt{b_1}$ or b_2. The parameters γ, δ are found by equating $\sqrt{\beta_1}$, β_2 of x to those of $\zeta + \lambda \sinh\left(\frac{X-\gamma}{\delta}\right)$ where X is a standard normal variate; this reduces to evaluating integrals of the form

$$(2\pi)^{-1/2} \int_{-\infty}^{\infty} \exp(-1/2\ y^2)\ f(y)^s\ dy, \qquad f(y) = e^{y/\delta},\ s = 1,2,\ldots .$$

For ζ, λ we use

$$E(x - \zeta)/\lambda = -\omega^{1/2} \sinh\Omega ,$$

$$\text{Var } (x - \zeta)/\lambda = 1/2(\omega - 1) \quad [\omega\cosh (2\Omega) + 1]$$

where $\omega = \exp(1/\delta^2)$, $\Omega = \gamma/\delta$.

It should be clearly understood that X is an equivalent normal deviate for say $\sqrt{b_1}$, but only as far as the first four moments of $\sqrt{b_1}$ are taken into account. In general, S_U does a better approximation to $\sqrt{b_1}$ than to b_2 (Shenton and Bowman, 1975).

A natural projection from S_U approximations was made by D'Agostino and Pearson (1973). They suggested that in normal sampling

$$\chi^2 = \chi^2(\sqrt{b_1}) + \chi^2(b_2) \tag{5}$$

is approximately distributed as χ^2 with $\nu = 2$ degrees of freedom. However, they had overlooked the fact of dependence which was pointed out by Anscombe (D'Agostino and Pearson, 1974).

Bowman and Shenton (1975) derived contours at various probability levels for accepting normality based on a modified form of equation (5). They used

$$Y_S^2 = \chi^2(\sqrt{b_1}) + \chi^2(b_2) \tag{6}$$

without the χ^2 assumption, finding the 5%, 10%,..., 90%, 95% values by large-scale simulation, and mapping the relation back onto the $\sqrt{b_1}$, b_2 plane. A scrutiny of the contours (Bowman and Shenton, 1975, p. 247) brings out a discrepancy between the contours and the Monte-Carlo dot shape; the contours are flattened ellipses, whereas the dot outline appears more akin to concentric parabolic arcs. Another aspect of this is to say that the $\sqrt{b_1}$ arrays become multimodal as b_2 increases. (In the absence of this diagram, readers will find a similar phenomen displayed in Figure 1-e).

A parallel study by Tietjen and Lowe (1975) displays several three-dimensional plots of the distribution of $\sqrt{b_1}$, b_2 along with a set of contours at the 95% level, sample sizes ranging from 4 to 50, the results

being mainly based on Monte-Carlo simulations. Their contours do have shapes similar to concentric parabolic arcs, although the gradient change is sharp at the intersections.

CONTOURS IN NON-NORMAL SAMPLING

The Johnson S_U approximations for $\sqrt{b_1}$, b_2, namely

$$X_1 = X(\sqrt{b_1}) = \gamma_1 + \delta_1 \sinh^{-1} (\sqrt{b_1} - \zeta_1)/\lambda_1$$
$$X_2 = X(b_2) = \gamma_2 + \delta_2 \sinh^{-1} (b_2 - \zeta_2)/\lambda_2 \qquad (7)$$

can be set up in non-normal sampling--actually the first four asymptotic moments of $\sqrt{b_1}$, b_2 are programmed (using terms through order n^{-3}) followed by an iterative scheme to determine the parameters in equation (7).

There will now be correlations between $\sqrt{b_1}$, b_2 and X_1, X_2; define R = correlation (X_1, X_2). To construct contours in this case, replace equation (6) by the approximate χ^2 variate

$$Y_S^2 = (X_1^2 - 2RX_1X_2 + X_2^2)/(1 - R^2) \qquad (8)$$

where R is intractable mathematically. For a specified population (Pearson Type I, normal mixture) Monte-Carlo assessments of R can be found from equation (7). The results for 50,000 samples are shown in Table 2 along with other moments of X_1, X_2. The stability of R for variable sample sizes is noteworthy as well as its robustness with respect to populations. Having assessed R, Monte-Carlo assessments of the percentage points of Y_S^2 from equation (8) could be found by a second simulation run in each case; however, since our interest in equation (8) is to reach a broad understanding as to the effect of skewness in the parent population on the shape and spread of Y_S^2 contours, we have used Y_S^2 on the assumption that it is approximately distributed as χ^2 with two degrees of freedom. The populations sampled are given in Table 3, and the contours in Figures 1-a to 1-e.

TABLE 2. Mean, Standard Deviation, Correlation for $X_S(\sqrt{b_1})$, $X_S(b_2)$ in Sampling from Pearson Type 1 ($\sqrt{\beta_1}$ = 2/7, β_2 = 33/14) and a Normal Mixture for Various Sample Sizes

n		$E(b_2\sqrt{b_1})$	ν_{10}	σ_{10}	ν_{01}	σ_{01}	R
30	(a)*	0.499	-0.006	0.996	-0.004	0.986	0.508
	(b)	0.443	0.010	1.001	0.004	1.001	0.442
50	(a)	0.551	0.003	1.003	0.004	1.003	0.548
	(b)	0.459	0.002	0.994	-0.002	0.998	0.463
80	(a)	0.565	-0.009	0.996	-0.005	0.999	0.568
	(b)	0.466	0.003	1.000	0.006	0.995	0.468
100	(a)	0.567	0.002	0.998	0.003	0.997	0.571
	(b)	0.469	0.003	0.998	-0.006	0.999	0.470

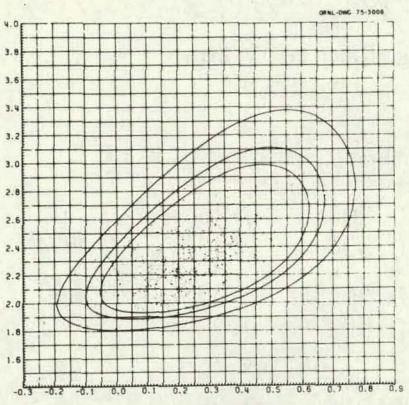
*Populations: (a) Type I, (b) Normal mixture with same $\sqrt{\beta_1}$, β_2
Moments based on 50,000 samples. Notation: $\nu_{10} = EX_1$, $\sigma_{10} = \sqrt{\text{Var}}\ X_1$, $\nu_{01} = EX_2$, $\sigma_{01} = \text{Var}\ X_2$, R = correlation (X_1, X_2).

TABLE 3. Populations Used in Y_S^2 Contours

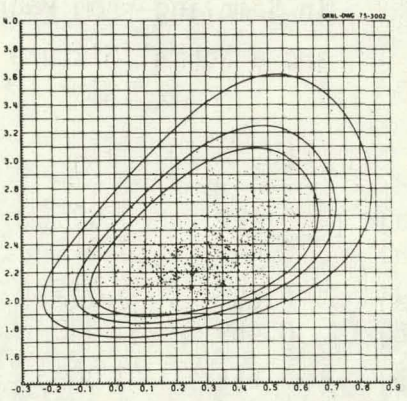
Figure	Population	$\sqrt{\beta_1}$	β_2	n	R	Number of $(\sqrt{b_1}, b_2)$ points in 1000 outside the contour 90%	95%	99%
1a	Type I*	2/7	33/14	100	0.571	89	46	9
1b	N.M.	2/7	33/14	100	0.470	85	41	6
1c	N.M.	1.0	2.2	70	0.998	67	37	16
1d	N.M.	1.0	2.6	50	0.979	97	48	13
1e	Uniform	0.0	1.8	50	0.000	100	57	16

*Type I has indices 2 and 3; N.M. = Normal Mixture.

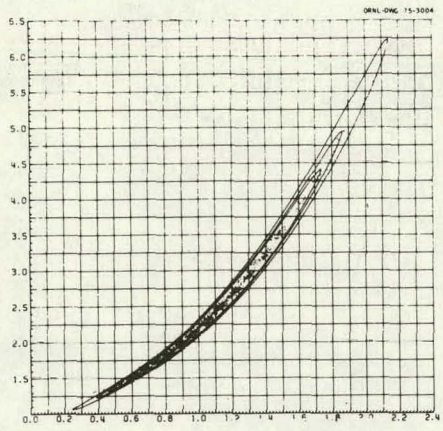
VERTICAL AXIS: b_2 HORIZONTAL AXIS: $\sqrt{b_1}$



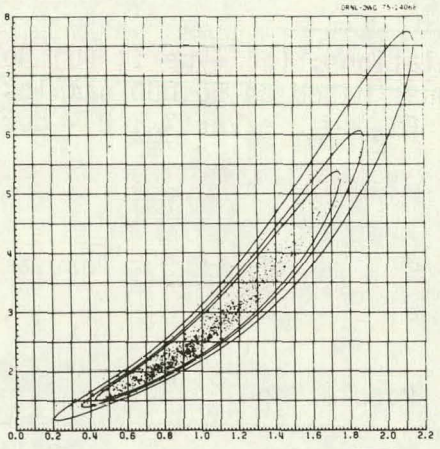
A. TYPE I; $\sqrt{\beta_1}=2/7$; $\beta_2=33/14$; n=100



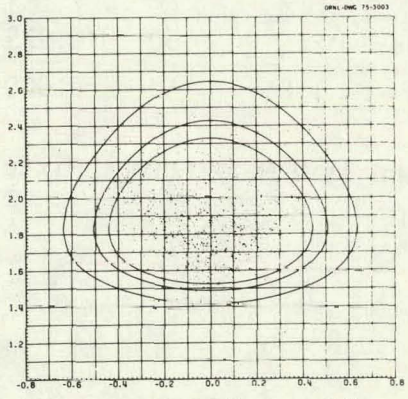
B. NORMAL MIXTURE $\sqrt{\beta_1}=2/7$; $\beta_2=33/14$; n=100



C. NORMAL MIXIURE; $\sqrt{\beta_1}=1$; $\beta_2=2.2$; n=70



D. NORMAL MIXTURE, $\sqrt{\beta_1}=1$; $\beta_2=2.6$; n=50



E. UNIFORM, n=50

FIGURE 1. Y_S^2 Contours at 90,95 and 99%

REMARKS ON THE CONTOURS

There is little to choose between the contours in Figures 1-a and 1-b which refer to two populations with the same first four moments.

When $\sqrt{\beta_1}$ is large and β_2 is in the neighborhood of $1 + \beta_1$ (the critical boundary, since in general $\beta_2 > 1 + \beta_1$), then R is near to unity and the regions are "narrow" and "long"; as R decreases, the regions fill out rapidly (Figures 1-c and 1-d) and become more box-shaped, as would be expected for independence.

JOINT DISTRIBUTION OF $\sqrt{b_1}$, b_2

PEARSON AND PRETORIUS

At the outset something needs to be said about the order of difficulty that confronts those who attempt the resolution of bivariate, trivariate and higher distributions. One may recall the great disparity between studies of univariate and multivariate situations. For example, there are scores of univariate discrete models (Poisson, binomial, Neyman's A, B, C, etc.) but rare cases of bivariate studies and still rarer studies of trivariate. Indeed bivariate discrete cases frequently revolve around the Poisson and negative binomial models and are thus fairly obvious "extensions."

Karl Pearson in a series of papers starting around 1905 made an heroic effort (1925) to fit a bivariate surface using two means, three covariances, four third-order and five fourth-order moments; the paper was called "The Fifteen Constant Bivariate Frequency Surface." He was apparently convinced that no satisfactory results would accrue for anything less than a four-moment bivariate fit, and actually tried what we should now identify as a Gram-Charlier series based on the bivariate normal as kernel. He was not enchanted with the result (p. 269), and one would gather did not expect good fits to result (as measured by χ^2).

Another landmark in the bivariate situation is the work of Pretorius (1930), who was certainly not hampered by small samples; one example has n = 631,682. He studied several bivariate surfaces, derived regression curves, marginal distributions and set up contours of equal probability.

The present state of the subject is indicated by the scope of the recent text by Johnson and Kotz (1972). It seems to us, that there are still very few multivariate cases in which surfaces are successfully fitted to empirical data, and that, in general, much less is to be expected once one crosses the univariate boundary.

A MODEL FOR $\sqrt{b_1}$, b_2

Our original aim was to consider samples from general Pearson populations, but then it was realized that even the normal case has not been resolved. The extensive study of the simulated ($\sqrt{b_1}$, b_2) couplet diagrams was certainly encouraged by much helpful correspondence with E. S. Pearson. Let it be said, however, that his interest was in the power of various tests (Shapiro-Wilks, W', etc.), and we became intrigued by the problem of finding a bivariate model which at least has the "right" shape for its equi-probability contours. Monte-Carlo simulations (samples of 30, 50 mainly) were made of from 50,000 to 100,000 runs, the couplets being set out on a two-dimensional grid of values for $\sqrt{b_1} = 0(\pm 0.05) \pm 2$, $b_2 = 1(0.05)7.0$. It was disappointing to notice quite frequent cases of lack of smoothness in the arrays, even the array for $|\sqrt{b_1}| < 0.1$ being somewhat irregular. Nonetheless certain broad features were evident. The arrays for $\sqrt{b_1}$ constant were of the form of Type III (gamma), bounded very nearly by the parabolic arc $b_2 = 1 + b_1$, while the array means followed a parallel parabolic arc. The arrays for b_2 constant were more problematical, and became bimodal as b_2 increased.

NEW BIVARIATE MODEL

The model finally used was multiplicative, with

$$f(\sqrt{b_1}, b_2) = w(\sqrt{b_1})\ g(b_2|\sqrt{b_1}) \qquad (9)$$

where

$$w(\sqrt{b_1}) = \frac{\delta}{\sqrt{2\pi}\ [\lambda^2 + (\sqrt{b_1} - \xi)^2]}\ \exp\ \{-\tfrac{1}{2}[\gamma + \delta\ \sinh^{-1}(\frac{\sqrt{b_1} - \xi}{\lambda})]^2\}$$

is Johnson's S_U approximation to the distribution of b_1 and function g is a conditional gamma distribution with

$$g(b_2|\sqrt{b_1}) = \frac{k[k(b_2 - 1 - b_1)]^{p(\sqrt{b_1}) - 1}}{\Gamma[p(\sqrt{b_1})]} \exp[-k(b_2 - 1 - b_1)]$$

and $k > 0$, $p(x)$ positive for all real x. The bivariate density equation (9) may be regarded as a crude approximation to the distribution of $(\sqrt{b_1}, b_2)$ in general sampling.

A comparison of the 90, 95, 99% contours for the Y_S^2 model (see equation (8)) and new bivariate model equation (9) in samples of 50 from the normal is shown in Figure 2-a. To bring out the versatility of both models (Y_S^2 and $f(\sqrt{b_1}, b_2)$) the contours are also given for samples of 50 from the normal mixture for which $\sqrt{\beta_1} = 1.0$, $\beta_2 = 2.6$ (two components, equal variances), see Figures 1-d and 2-b. The fitting of the bivariate model is discussed elsewhere.

As one simple approach to the goodness-of-fit of the models, we have drawn one thousand couplets $(\sqrt{b_1}, b_2)$ for each of these examples. For the normal case there were 98, 59, 10 points outside the 90, 95, 99% contours in good agreement, Similarly, the corresponding results for the normal mixture were 84, 48, 17.

CONCLUSIONS

We have presented some of the problems relating to the distribution of the $\sqrt{b_1}$, b_2. A great deal of effort has gone into the problems associated with the marginal distribution in the normal case during the past half century. Our present work has produced two simple bivariate models in general sampling.

At least the construction of the models has highlighted the main difficulties. However, it should be noted that in the normal case there is an obvious null hypothesis for the departure tests, whereas this aspect runs into conceptual problems in the non-normal case.

VERTICAL AXIS: b_2 HORIZONTAL AXIS: $\sqrt{b_1}$

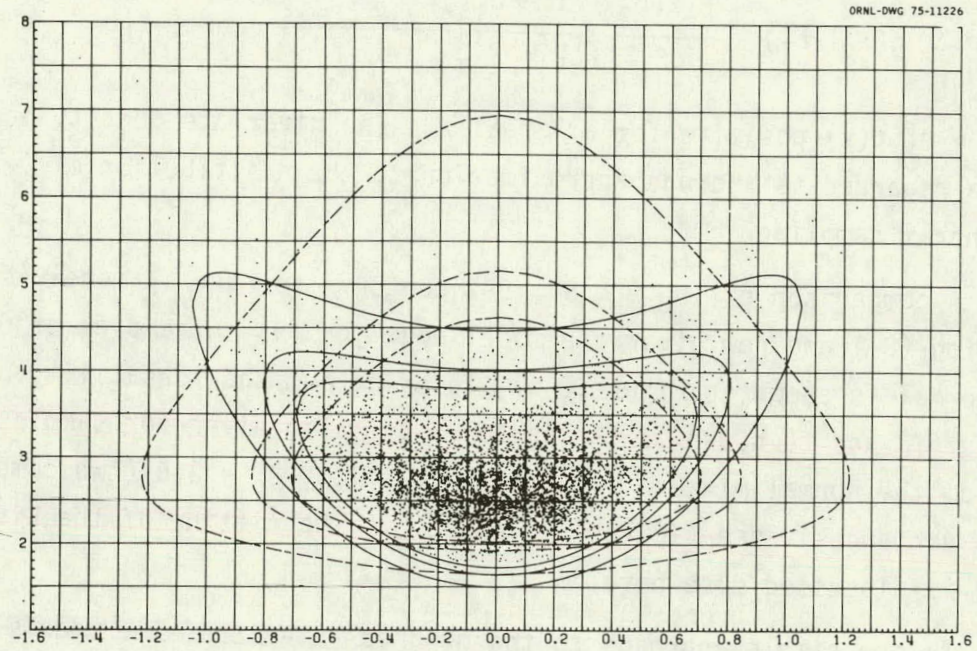


FIGURE 2-a. Y_S^2 and BiVariate Contours at 90,95 and 99%

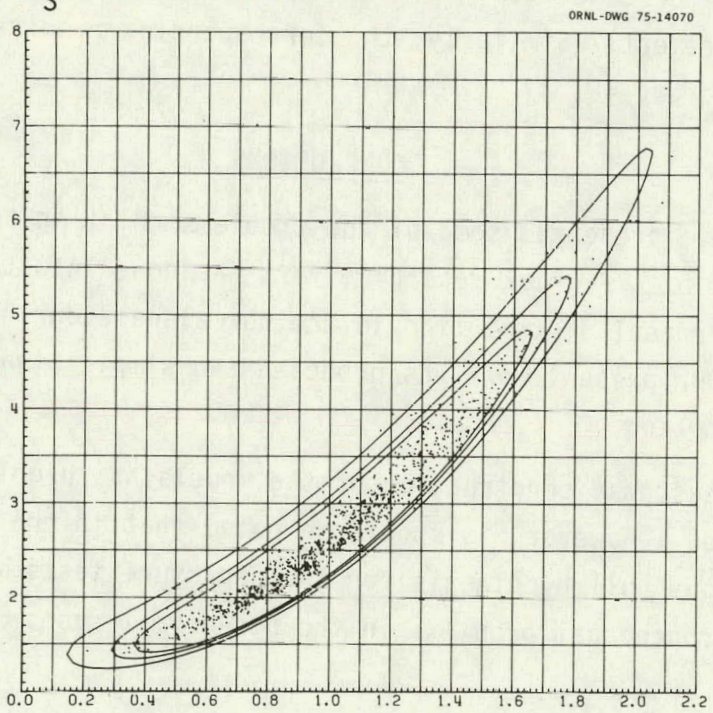


FIGURE 2-b. BiVariate Contours at 90,95 and 99%

REFERENCES

Anscombe, Francis J., and Glynn, William J., "Distribution of the Kurtosis Statistic b_2 for Normal Samples," Technical Report No. 37, Yale University, Department of Statistics, 1975.

Baker, George, A., Jr., Essentials of Pade´ Approximants, Academic Press, New York, 1975.

Baker, George A., Jr., and Gammel, John L., (Editors), The Pade´ Approximant in Theoretical Physics, Academic Press, New York, 1970.

Bowman, K. O., and Shenton, L. R., "Tables of Moments of the Skewness and Kurtosis Statistics in Non-normal Sampling," Union Carbide Nuclear Division Report UCCND-CSD-8, 1975.

Bowman, K. O., and Shenton, L. R., "Omnibus Test Contours for Departures from Normality Based on $\sqrt{b_1}$ and b_2," Biometrika, 62 (1975), 243-50.

D'Agostino, R. B., "Transformations to Normality of the Null Distribution of g_1," Biometrika, 57 (1970), 679-81.

D'Agostino, R. B., and Pearson, E. S., "Tests for Departures from Normality. Empirical Results for the Distribution of b_2 and $\sqrt{b_1}$," Biometrika, 60 (1973), 613-22.

D'Agostino, R. B., and Pearson, E. S., "Correction and Amendment. Tests for Departures From Normality. Empirical Results for the Distribution of b_2 and $\sqrt{b_1}$," Biometrika, 61 (1974), 647.

Fisher, R. A., "Moments and Product Moments of Sampling Distributions," Proc. Lond. Math. Soc. (Series 2), 30 (1928), 199-238.

Fisher, R. A., "The Moment of the Distribution for Normal Samples of Measures of Departures from Normality," Proc. Royal Society, Series A, 130 (1930), 17-28.

Geary, R. C., and Worlledge, J. P. G., "On the Computation of Universal Moments of Tests of Statistical Normality Derived from Samples Drawn at Random from a Normal Universe. Application to the Calculation of the Seventh Moment of b_2." Biometrika, 34 (1947), 98-110.

Hsu, C. T., and Lawley, D. N., "The Derivation of the Fifth and Sixth Moments of b_2 in Samples from a Normal Population," Biometrika, 31 (1939), 238-248.

Johnson, Norman L., and Kotz, Samual, Distributions in Statistics - Continuous Univariate Distributions, Vol. 1. Boston: Houghton Mifflin, 1970.

Johnson, Norman L., and Kotz, Samuel, Distributions in Statistics - Continuous Multivariate Distributions, New York: John Wiley, 1972.

Pearson, E. S., "A Further Development of Tests for Normality," Biometrika, 22 (1930), 239-249.

Pearson, E. S., "Some Problems Arising in Approximating to Probability Distributions, Using Moments," Biometrika, 50 (1963), 95-111.

Pearson, E. S., "Tables of Percentage Points of $\sqrt{b_1}$ and b_2 in Normal Samples: A Rounding Off," Biometrika, 52 (1965), 282-285.

Pearson, Karl, "On the General Theory of Skew Correlations and Non-linear Regression," Draper's Company Memoirs, Biometric Series II, 1905.

Pearson, Karl, "The Fifteen Constant Bivariate Frequency Surface," Biometrika, 17 (1925), 268-313.

Pepper, J., "Sampling of Third Moment Coefficient," Biometrika, 24 (1932), 55-84.

Pretorius, S. J., "Skew Bivariate Frequency Surfaces, Examined in the Light of Numerical Illustrations," Biometrika, 22(1930), 109-223.

Shenton, L. R., and Bowman, K. O., "Johnson's S_U and the Kurtosis Statistics," Journal of the American Statistical Association, 70 (1975) 220-28.

Shenton, L. R., Bowman, K. O., and Sheehan, D., "Sampling Moments of Moments Associated with Univariate Distributions," J. Royal Statistical Soc., Series B., 33, 3, (1971) 444-57.

Tietjen, Gary L., and Lowe, Victor W., Jr., "Some Aspects of the Joint Distribution of Skewness and Kurtosis in Normal Samples," (private communication).

Wishart, J., "The Deviation of Certain High-Order Sampling Product-Moments from a Normal Population," Biometrika, 22 (1930), 224.

Van Dyke, Milton, "Analysis and Improvement of Perturbation Series," Quarterly Journal of Mechanics and Applied Mathematics, 27 (1974) 424-50.

DISCUSSION OF THE PAPER BY BOWMAN AND SHENTON, "DISTRIBUTION OF SKEWNESS AND KURTOSIS STATISTICS"

G. L. Tietjen
Los Alamos Scientific Laboratory
Los Alamos, New Mexico

This paper contains a good historical introduction to the statistics $\sqrt{b_1}$ and b_2. These statistics have been widely used (1) to characterize the shape of a given distribution, (2) as tests of normality or the lack of it, and (3) as tests for outliers from a normal distribution, particularly for the case of several outliers. It seems likely to me that they will find an even wider use in tests of goodness-of-fit for distributions other than the normal and in tests for outliers from distributions other than the normal. Recently, Bob Hogg has discussed adaptive robust estimation. In particular, he has recommended estimators of central tendency (mean, median, etc.) which depend on some vague notion of the shape of the distribution. For one shape, the mean is recommended. For another shape, he wants us to use the median, etc. I foresee that $\sqrt{b_1}$ and b_2 will be used for this adaptive estimation.

In view of all this, isn't it surprising how little we know about the distribution of $\sqrt{b_1}$ and b_2? People like Pearson, Fisher, Wishart, and Geary have spent a good deal of time on them, not to mention Bowman, Shenton, D'Agostino, and Anscombe. I first became interested in this area when I saw Ferguson's table of critical values of $\sqrt{b_1}$ and b_2 which he used to detect outliers. It had such gaps in it that I did a Monte Carlo simulation in which I calculated, for normal samples, the two statistics. I then ordered the values of $\sqrt{b_1}$ and b_2 and picked off the percentiles. I sent this to Biometrika. The editor immediately replied that he had just accepted a very similar paper by Ralph D'Agostino on that subject. Could the two of us collaborate to produce a joint paper? We did write two papers together, and I would still like to meet D'Agostino some day. I tell this to show you how little was known about the percentiles of these statistics, just five years ago for small sample sizes.

Kim Bowman has given an account of the approaches which have been made toward approximating these distributions. There were the series expansions in k-statistics. That, thank heavens, never caught on with the Americans. Then the moments (up to the 8th) were developed, and expressed in series in powers of 1/n. These were used to fit the Pearson and the Johnson systems of distributions. Reasonable fits for $\sqrt{b_1}$ were made for $n \geq 8$, but samples sizes of at least 40 had to be used to get good approximations for b_2. Also, using these moments, D'Agostino found that standardizing $\sqrt{b_1}$ (by subtracting the mean and dividing by the standard deviation) would do a pretty good job of transforming $\sqrt{b_1}$ to normality.

Shenton and Bowman used moments of the sample moments to approximate the distributions. These came out as Taylor's Series in powers of 1/n. In doing so, they had made much use of recursive relationships between the moments which quickly grew very complicated. Only Bowman's persistence at the computer made possible a continuation of their efforts. Since they were using the sample moments, they could work with whatever distribution they chose, and they turned out a lot of results, even generalizing to the bivariate case.

At about this point Victor Lowe and I became interested in using $\sqrt{b_1}$ and b_2 jointly as a test for normality. The big competitor was the W-test by Wilk and Shapiro. Looking at the simulation results with which they justified the omnibus properties of their test, it was evident that in most of the cases, either $\sqrt{b_1}$ or b_2 was more powerful than their test. The W-test was chosen as best because it did better over more alternatives than did the other statistics; not because it performed better with each alternative. So, we thought, a joint test may pick up the desirable features of both $\sqrt{b_1}$ and b_2. Perhaps it would do well where $\sqrt{b_1}$ did well and where b_2 did well.

To test our theory, we simulated 30,000 normal samples and calculated $\sqrt{b_1}$ and b_2 from each sample. Using $\sqrt{b_1}$ as one coordinate and b_2 as another, we plotted these--or rather we counted how many fell into each bin on a 100 x 100 array. The result was a bivariate, three-dimensional

surface which is the empirical joint density of $\sqrt{b_1}$ and b_2 (Figures 1 and 3). On the surface for n = 10, the sharp rise in the rear and the empty space in front are noted. These statistics are uncorrelated but dependent. If they were independent, the joint would be product of the marginals and would look like this Figure 2.

We were partly into the project when we noticed an article by D'Agostino and Pearson. They had standardized both $\sqrt{b_1}$ and b_2, then taken the sum of squares of these two variates which were approximately normal, and used a χ^2 distribution with two degrees of freedom to draw the contours of the joint distribution. We thought we were wrecked, but decided to continue, since they had said nothing about power studies on tests of normality. Just then Frank Anscombe wrecked their theory by pointing out that $\sqrt{b_1}$ and b_2 were not independent, hence you couldn't use the χ^2 theory.

We then tried to devise contours which were functions of the sample size n, and the best we could do was to intersect two parabolas. This crescent-shaped contour was too sharp at the tips.

We finished and sent the paper off to Biometrika. Afterwards, Kathy Campbell did some Fourier smoothing for us and produced 90, 95 and 99% contours from 100 x 100 arrays (Figure 4). These are contours of equal elevation. We would like to recommend our techniques for dealing with bivariate densities in general. Then a letter came from the editor saying that they had accepted a very similar paper by Bowman and Shenton, but would we get together and talk anyway? None of us knew quite what that meant, but we did talk, and some of this today has been an outgrowth of our discussions.

What had Bowman and Shenton done? They had used Johnson's transformations to normality, involving the arcsinh. Their contours still resembled the case of independence. D'Agostino's contours looked like theirs, but his did not have the right content, whereas theirs did. Nevertheless, they realized that their contours did not have the right shape. They were not contours of equal elevation on the surface.

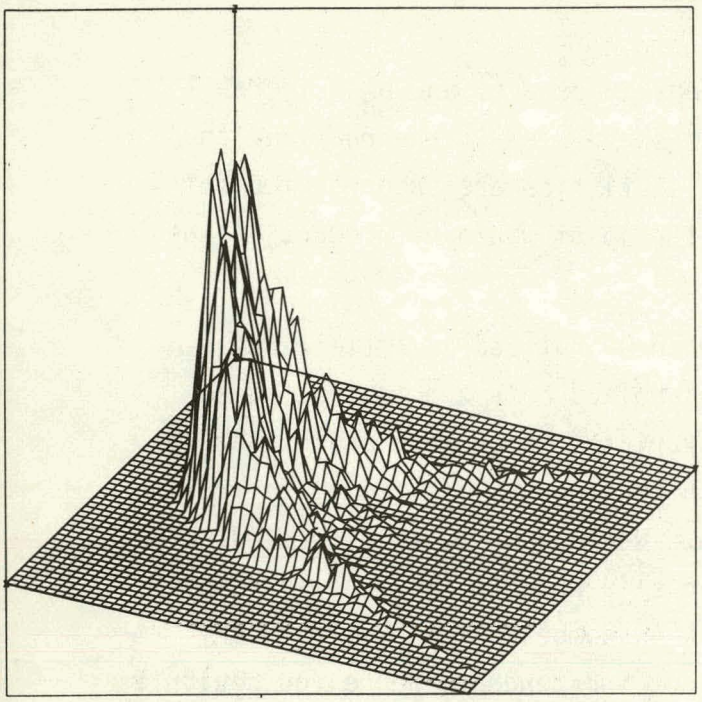


FIGURE 1. Joint Density of $\sqrt{b_1}$ and b_2, Normal Sampling, N=10.

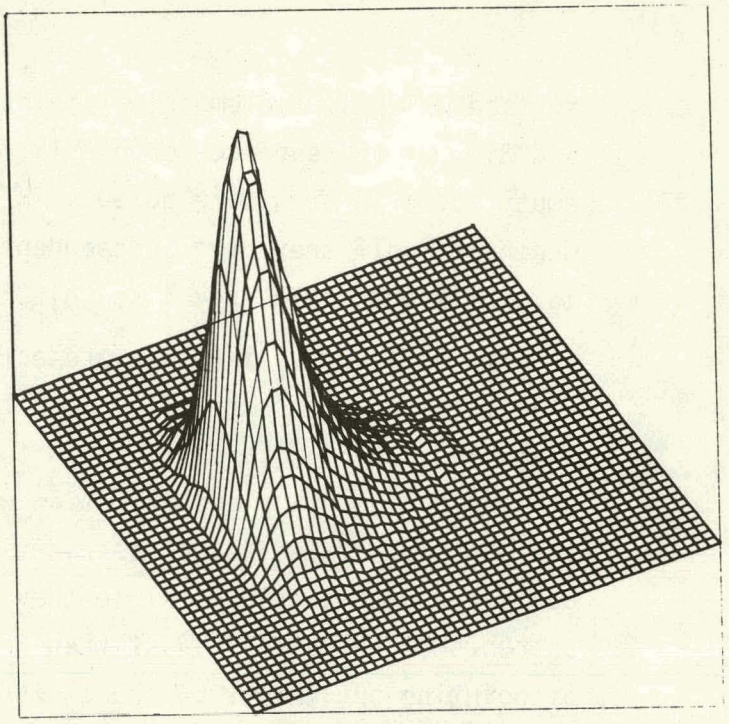


FIGURE 2. Joint Density Under Assumption of Independence, N=10.

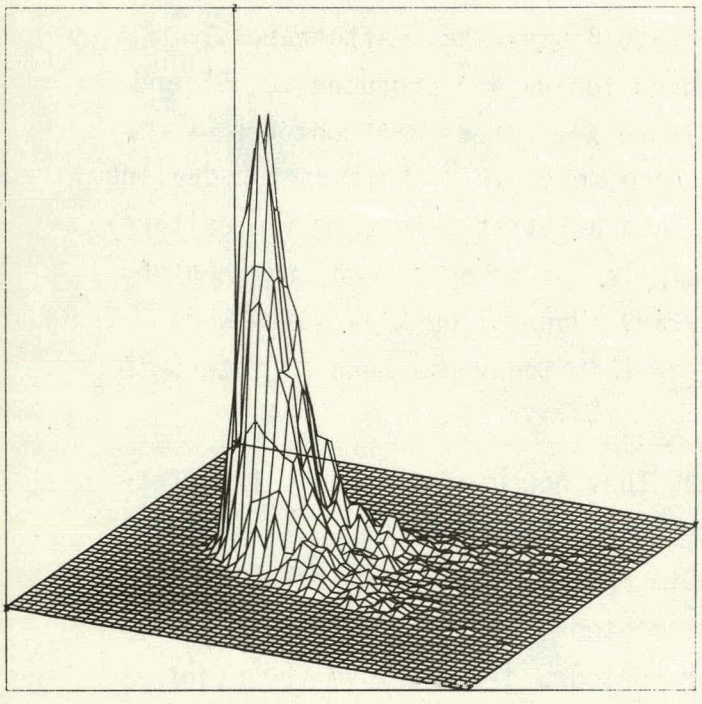


FIGURE 3. Joint Density for N=40 Effect of Increasing N.

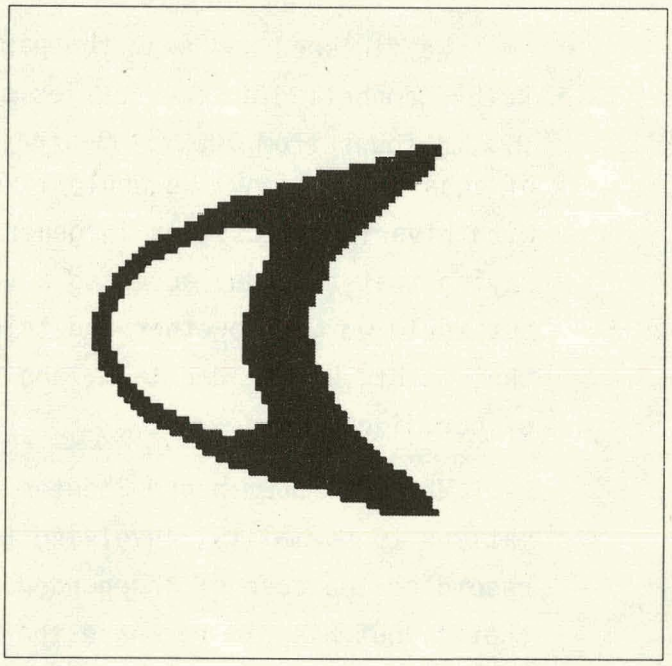


FIGURE 4. 90, 95, and 99% Contour Intervals, N=10.

In a second try, they came up with what they have called a crude approximation to the joint density of $\sqrt{b_1}$ and b_2. It is the product of Johnson's S_u approximation for $\sqrt{b_1}$ and another function involving both $\sqrt{b_1}$ and b_2. The latter function is a <u>conditional</u> gamma distribution. The shape of their contours is now much better, and the content is correct. This aspect of the problem thus appears to have been solved. With the contours which Victor and I came up with we could not quite match the power of the W-test. We await anxiously Kim's Monte-Carlo investigation of power on her contours. Possibly this may revolutionize tests of normality.

PROBLEM PRESENTATIONS AND DISCUSSIONS

HOW SHOULD THE LOSS OF COOLANT ACCIDENT BE STUDIED? SOME BACKGROUND AND PERSPECTIVE

R. G. Easterling
Nuclear Regulatory Commission
Washington, DC

The total problem in the study of LOCA (loss of coolant accident), as I understand it, is the following: How does one obtain and use information from less than full-scale experiments and from computer models to make reliable predictions of what would happen to a reactor in the event of a LOCA? Stating the problem simply does not mean the problem is simply solved. There are many sub-problems and this and the following presentations will concentrate on but a few of them.

As a framework for discussion, consider a general problem, expressed as follows: There is a characteristic of interest, Y, which is a function of several arguments. For various reasons there is "uncertainty" about the arguments. The problem is how to translate that uncertainty into uncertainty about Y. Symbolically, let the relationship between Y and the arguments of the function be denoted by:

$$Y = f(X_1, X_2, \ldots, X_m; Z_1, Z_2, \ldots, Z_n),$$

where the two types of arguments are used to indicate two different sorts of uncertainty associated with the arguments. These are:

1. nonconstant physical characteristics: $X_1, X_2, \ldots, X_m$;
2. the use of estimated constants: $Z_1, Z_2, \ldots, Z_n$.

For example, suppose Y is a characteristic of a reactor at some randomly selected time during its life. That characteristic might depend on the power level at which the reactor is operating, say X_1, and so the uncertainty in this case is that associated with the presence of a random variable. Not all nonconstant physical characteristics are random variables. For example, in another problem, X_1 might be the inlet water temperature which is known to follow a specific seasonal pattern. What one might be interested in is the resultant pattern in Y, not in what Y might be for a randomly selected season.

The second type of uncertainty is that attributable to using estimates of constants. For example, Y might be the reliability of a system and Z_1 the reliability of a component in that system. Not knowing Z_1, one might replace it by an estimate, and that estimate might be based at least in part on data. The X's and the Z's might also be related in that the probability distribution of some X's might be known except for some constants which must then be estimated.

Perhaps other sorts of uncertainty could be distinguished, but these two are ample for the purposes of illustration and the following presentations. Other uncertainties, such as how well the model agrees with reality must ultimately be considered but are not the topic here. Note also the following distinction: Statisticians are used to seeing the uncertainty associated with an equation, such as that given in the opening paragraph, designated by ϵ added to the end of the equation. Here the uncertainty is inside the function and is not so simply expressed. In the following paragraphs, the problem mentioned refers to that set forth on the previous page.

Two quite visible instances in which this problem occurs in the nuclear industry are the Reactor Safety Study (1974) and the analysis of Loss of Coolant Accident (LOCA) computer codes. In the first, Y might be the probability of exposing 1000 people to a specified amount of radiation in the event of a nuclear reactor accident. The arguments would include the reliabilities of the various safety systems, meteorological conditions, and population distribution. In the second instance, the LOCA computer code simulates an accident and yields a record of cladding temperature as a function of time. The peak of that curve is one characteristic of interest and the arguments of the function (the computer code is the function) are numerous and fall in both categories given above.

The following presentations will examine that example in more detail. As a preliminary, let us consider the statistical aspects of the general problem.

Is it a statistical problem? Perhaps. The uncertainty in some instances may be traced to an identifiable set of data. Also, as we shall see, where the problem involves how to

choose points at which to evaluate the function, statistical experimental design ideas can be used.

Is the problem usually approached from a statistical point of view? I say no. It is amazing to me how quickly engineers (or physicists, or chemists, or some statisticians) can go from a vague expression of uncertainty to a completely specified probability distribution. (There is some sort of reverse Central Limit Theorem at work here. The less data, the easier it is to specify a probability distribution.) Thus the problem becomes one of applied probability, not statistics, a distinction I think we should make.

Is the problem one that statisticians can profitably tackle? I think so. The problem is essentially one of information processing: How does one merge the information about the arguments to yield information about Y? I think information processing (as opposed to statistical inference or decision making) is primarily what statisticians do and should do. In this particular regard, statisticians should be particularly sensitive to the insertion of artificial information, such as a completely specified distribution, which could distort or even determine what one concludes from the analysis.

What work has been done on the problem by statisticians? Little in general, I think, but there are some special cases which have been studied. One particular instance is the following: The Z's are component reliabilities, Y the system reliability, and the uncertainty about the Z's is due to the fact that all that is known about them is that n_i of component i have been tested and f_i failures occured (the failure information could include times to failure). In general, we must use approximate methods to process this information into confidence (or consonance, see Kempthorne and Folks, 1971) limits on Y and only in recent years with work by Nancy Mann and co-authors (1974) have we had generally usable and accurate approximations.

In another class of situations which includes statistical tolerancing, all of the arguments are random variables. All that is known about their distributions is the information contained in samples from each. It is my impression that statisticians have not progressed beyond point estimation in this case. We find by some means the "best" fitting distribution for the arguments and then by analysis, error propogation, or Monte Carlo, depending on the complexity of the function, obtain the derived distribution of Y. I am not aware of any published efforts to quantify the uncertainty associated with the derived distribution, such as by obtaining statistical tolerance intervals, but it seems worthwhile doing. (Note that statistical tolerance is used in two distinct ways in this paragraph--one of our problems in communicating with nonstatisticians.)

My summary point is that we statisticians are still at the rudimentary stage in the development of methods to handle this problem. And it is an important problem. Millions of dollars can be riding on the outcomes of these functions. Further, as I imagine many readers can attest, it is a problem which arises repeatedly in the nuclear industry and others. While I may not always agree with whatever solution is presented, (the usual approach seems to be to calculate the root sum of squares of something), I regard it as a healthy sign that people who work with these functions recognize the existence of uncertainty in them and the need to evaluate that uncertainty. We should encourage them and help find correct and credible ways to accomplish this.

REFERENCES

Reactor Safety Study, Draft, U.S.AEC WASH-1400, August 1974.

Kempthorne, O. and Folks, J. L., Probability, Statistics and Data Analysis, Iowa State University Press, 1971.

Mann, N. R., Schafer, R. E., and Sinpurwalla, N. D., Methods for Statistical Analysis of Reliability and Life Data, Wiley, 1974.

HOW SHOULD A LOSS OF COOLANT ACCIDENT BE STUDIED?

G. P. Steck
Sandia Laboratories
Albuquerque, New Mexico

INTRODUCTION

Easterling's paper discusses the general problems associated with studying the uncertainty generated by a function of many variables when there were attending uncertainties in those variables. McKay's paper discusses the problem of developing a computer code that models reality adequately. I will discuss a particular problem and in doing so assume that I have an adequate code. The problem I will describe is an important one to the nuclear field.

One of the misfortunes that is postulated in analysis of water reactors is loss of coolant through, for example, a ruptured pipe. According to the Rasmussen study, such a loss of coolant accident, or LOCA as it is called, is very unlikely. Nevertheless, it is receiving a great deal of attention because of the stringent regulatory requirements for emergency core cooling systems and analyses.

There has not been, and in the immediate future won't be, any full-scale test of an entire system--it would be prohibitively expensive--so the effort has been in scale testing and computer model building. Statistical investigations can help this effort through assessment of experimental and computer modeling information. I will discuss a portion of the latter subject.

For discussion purposes, I will assume there is a realistic model for a LOCA in the sense that there is one set of values (unknown) of the engineering parameters such that for the set of parameters defining any accident the computer code gives a suitably accurate LOCA time history. What kind of uncertainties will exist in modeling a LOCA? First, there are the uncertainties in the accident parameters themselves: system configuration, operating history, inlet water temperature, type of pipe break, etc. Second, there are the engineering "correlations" that exist by the hundreds in the submodels. These are actually estimated regression parameters and may be

based on data or engineering judgment. We wish to investigate the effect of these uncertainties upon the LOCA time history. However, the investigation is seriously complicated by two characteristics of the computer model:

1) It has many variables, possibly thousands. We want to treat as many as we can.
2) It can take as many as 8 or 10 hours of computer time to produce a LOCA history. One is strongly motivated to make as few evaluations of the function as possible.

The immediate problem is how to conduct meaningful LOCA investigations in light of these conflicting requirements.

The instructions I received concerning the preparation of papers for these Problems Sessions contained the injunction to devote as much time as possible to the description of the problem. This I think I have done. Now I will describe some possible solutions as I see them. Hopefully this will provide a starting place for the discussion that will follow.

MONTE CARLO SOLUTION

The conceptually simplest method is that of Monte Carlo. Using the distributions determined for the inputs, one would sample them and evaluate the function (the computer model).

Repeated evaluations could be utilized in statistical analyses of LOCA behavior. For example, one could study peak clad temperature, which is limited to no greater than 2200°F when analyzed in a conservative manner as specified by the Code of Federal Regulations. One could assess the conservatism by finding an upper confidence limit for some percent point and comparing that with 2200°F or by finding a lower confidence limit for $P(PCT \leq 2200°F)$ and seeing how large it is.

This would be a very satisfying procedure except for three things:

1) Where do the input distributions come from and what happens if they are wrong? One must consider the problem of sensitivity to the input distributions.
2) Suppose there are regions where the peak clad temperature is unbounded. Such regions with very high probability will very

likely be found; regions with very low probability can be ignored. What happens in between?

3) A fairly large number of function evaluations must be made to get satisfactory results and this may be too expensive. $T_{max} = \max(T_1, T_2, \ldots, T_n)$ is a 90% upper confidence limit for the 95th percentile for n = 45. A similar 90% upper confidence limit for the 99th percentile requires n = 230, and if $T_{max} \leq 2200$, a 100α percent lower confidence limit for P(PCT $\leq$ 2200°F) is $(1 - \alpha)^{1/n}$. To make $0.1^{1/n}$ equal to 0.99 requires n = 230. This is clearly unsatisfactory for 10 hours/evaluation.

My conclusion is, therefore, that a Monte Carlo solution would be nice--but.

RESPONSE SURFACE ANALYSIS

Another way of treating the problem is to use the ideas of response surface analysis. Here the goal is to fit the peak clad temperature surface by function evaluations at a sequence of carefully chosen points, probably a sequence of fractional factorials. Once a fit is obtained it can be exhaustively studied by Monte Carlo or by numerical integration, since its running time will be seconds instead of hours. The difficulty with this method, obviously, is getting a good fit along with some assurance that one has a good fit.

Obtaining a good fit is, for me at least, too much a question of artistry and not enough of science. Our *modus operandi* on some sample problems has been to continue evolving a model until new evaluations agree sufficiently well with model predictions to warrant stopping. If they don't, a new regression is done with all runs, old and new, to get a new model, and the evolution continues.

Another difficulty is that is not possible to make precise probability statements about how good the fit is. Since any quantification of the degree of conservatism should only refer to the fitted surface actually obtained, call it $\hat{T}(X)$ and not to any ensemble of possible surfaces, one is left with a precise estimate of $P[\hat{T}(X) \leq 2200°F]$ which is obtained by exhaustive sampling of a surface whose degree of fit is not precisely known.

These two methods, Monte Carlo and surface fitting, are, to my mind, the principal contenders for attacking the problem and they both have their advantages and disadvantages. Negative feelings about the surface fitting method can be partly overcome by a history of good performance on practice problems. Let me now describe briefly what happened on two practice problems.

PRACTICE PROBLEM 1

The first function on which we compared the Monte Carlo and response surface techniques is given below as equation (1). It should be emphasized that while Easterling and I were experimenting with it we did not know what it was.

$$\hat{Y} = \min\left(3500, \frac{25\log(1.3F)}{C-15} + \frac{.9Be^{.15A}}{|B-E+10|} + 12e^{D/10}\right) \qquad (1)$$

The mean, standard deviation, maximum and minimum of each of the six variables, A, B, ..., F, were specified. A total of 45 evaluations were made, three of which hit the stop at +3500. The final fit was

$$\hat{Y} = \min\Big(3500,\ 242 + 1.56D + .00391D^3 - 60.2F + 4.16F^2 - .101DF + .00256D^2F^2 - 3.93B + .448BE - 399e^{-F/2} + E\cdot\left[12.5F - .666F^2 + 13.3e^{-F/2}\right]\Big/(C-15)\Big) .$$

The degree of fit can be seen from Table 1 which gives some percent points of both the true function, Y, and the fit, $\hat{Y}$, as estimated by 100,000 Monte Carlo samples. Furthermore, 95% of the errors, $Y - \hat{Y}$, were in the range -70 to +40.

TABLE 1. Percent Points of the Distributions of Y and $\hat{Y}$ as Estimated by 100,000 Samples in Problem 1.

Percent	Y	$\hat{Y}$	Difference
10	-470	-459	-11
20	-222	-215	-7
30	-139	-134	-5
40	-87	-85	-2
50	3	-5	8
60	132	135	-3
70	187	190	-3
80	271	273	-2
90	514	510	4
95	1001	987	14

For the Monte Carlo comparison we ran 10,000 runs of 45 each to obtain distributions for the percent points. For example, the median of 10,000 95% points of a sample of 45 was 825 and the first and third quartiles were 500 and 1355, respectively. Furthermore, the 90% upper confidence limit for the 95 percent point exceeded 3000 more than half the time.

It seems to me that the response surface method did a better job on this problem than Monte Carlo. It found the pole at C = 15, which was in a high probability region, and although it missed the infinite ridge at E - B = 10, which was in a low probability region, the errors are suitably small. The Monte Carlo method, on the other hand, has a great deal of variability.

PRACTICE PROBLEM 2

The second problem we used for comparing the response surface and Monte Carlo methods was a nineteen variable code designed by Warren Lyon (Nuclear Regulatory Commission) to produce representative LOCA temperature histories quickly. For this problem the dependent variable was peak clad temperature and the fitting process stopped after 165 runs of which 28 were unbounded.

The degree of fit attained in this problem can be seen from Table 2 which gives some percent points of the true and fitted responses as estimated by 7000 and 70,000 runs respectively. Furthermore, very nearly 90% of the absolute differences, $|Y - \hat{Y}|$, are less than 300°F, and 95% are less than 465°F.

This time, though, from the 7000 runs on the true response we could only simulate 42 sets of 165 for the Monte Carlo comparison. In these samples, the first, second and third quartiles of the distribution of the 95% point estimator were 1790, 1810 and 1840, respectively. Similarly, the first, second and third quartiles of the distribution of the 99% point estimator were 1910, ∞ and ∞, respectively. Thus, it seems that, on this problem, $\hat{Y}$ overestimates and Monte Carlo underestimates.

TABLE 2. Percent Points of the Distributions of Y and $\hat{Y}$ as Estimated by 7000 and 70,000 samples, respectively, in Problem 2.

Percent	Y	$\hat{Y}$	Difference
10	866	866	0
20	939	966	-27
30	1000	1047	-47
40	1065	1125	-60
50	1141	1205	-64
60	1238	1290	-52
70	1344	1390	-46
80	1495	1524	-29
90	1710	1748	-38
95	1846	1984	-138
98	1938	∞	
99	∞	∞	

SUMMARY

Let me now summarize what I have presented. First, what is my problem? My problem is how to find the distribution of a random variable that is a function of many other random variables, say 20 to 50, when the underlying distributions are known with varying degrees of precision and when each functional evaluation is expensive.

Second, what are some possible methods of attack? There are several and I discussed two in particular. These were simulation and surface fitting. I hope you can suggest others we haven't thought of.

Third, how do simulation and surface fitting perform on sample problems? It is our conclusion that surface fitting is a promising method of attack. It worked well on two sample problems and has the distinct advantage of being adaptive.

Finally, for the benefit of those who want more specific problems let me offer a few.

- Can the basic problem be approached in such a way that probability statements can be made about the degree of fit in surface fitting? Perhaps a hybrid of simulation and surface fitting is appropriate.
- It is time to make the stopping rule more precise. What are good stopping rules for surface fitting?

- What is the effect of sequential fractional factorial experimentation when the factor levels are adaptive?
- For validation purposes it would be nice to know the distribution of $\int_0^\infty [x(t) - \mu(t)]^2 dt$ for some suitably large class of stationary processes where $EX(t) = \mu(t)$, $t \geq 0$.
- What are the confidence limits for $\mu(t)$ for this class of nonstationary processes?

ON THE STATISTICAL ANALYSIS OF AN OUTPUT OF A LARGE COMPUTER CODE AS A FUNCTION OF THE INPUTS

M. D. McKay
L. A. Bruckner
Los Alamos Scientific Laboratory
Los Alamos, New Mexico

INTRODUCTION

The purpose of this presentation is to describe some statistical applications areas in our work related to the development and application of computer codes used by the United States Nuclear Regulatory Commission (NRC), Division of Water Reactor Safety Research. These codes, which employ state of the art techniques in hydrodynamics and thermodynamics, are used to study potential accidents and accident-related phenomena in nuclear reactors. The mathematical models in these codes are rarely perfect: either they depend upon quantities often imperfectly derived from experimental data or they employ approximations to not-quite-understood physics. Hence, the need for critical evaluations.

Throughout this dicussion the term output or response variable refers to a quantity calculated by the computer code. The term input or input variable is used in a very general sense to refer to noncalculated data or parameters. Also, the terms model and code are used interchangeably. A code, or model, is treated as a realization of a conceptual model.

We will look at some stages in modeling and how statistics might enter in, a classification of applicable statistical techniques, and examples of "sensitivity" analysis.

STAGES IN MODELING

To try to put what follows into perspective, I would like to consider some stages of a model's development: systems analysis, system synthesis, model verification, model validation, model analysis, and model application (Mihram, 1972). The first two, systems analysis and system synthesis, cover the initial study of the problem and the formulation of a structure, or computer code, which is assumed to adequately mimic reality. The third stage, model verification, is the determination of the correctness of the

model as compared to its intended structure. Model validation is a comparison of the output of the model with data. Model analysis consists of "shakedown" testing where the outputs, or responses, of the model are contrasted under various input conditions.

Model development is dynamic, passing back and forth through the stages and incorporating feedback until, finally, the model is accepted and is ready for application.

STATISTICS AND STAGES OF MODELING

MODEL VALIDATION

Where can, or should, statistical and probabilistic techniques come into play? I think the most obvious place is in model validation, the comparison of model response with experimental data. This application looks very much like the usual goodness of fit tests in regression problems. However, in depth considerations should be given to "error terms".

With regard to the experimental data, it is a good assumption that some error is involved. There are several possible sources of error. One occurs in the detection phase. As an example, suppose that an experiment is set up to determine an air flow in a pipe. In many cases the mere insertion of a flow meter into the pipe is enough to change the flow pattern so that valid determinations cannot be made. In this particular case, laser instrumentation seems to eliminate the problem, but the point remains. A second and obvious source of error is in the measurement of the response. Calibration error is an example. Finally, the process itself may be intrinsically stochastic: perfectly replicated experiments may produce different results because of the process itself.

Now let us look at sources of error related to both the experiment and the computer code. Initial conditions must be physically set up and/or determined before the experiment is carried out. The numerical values of these initial conditions are input to the code. Errors here mean that the code is not modeling the same problem that the experimenter is performing. A judgment of the validity of a code is difficult to make under these circumstances.

Lastly, with regard to the computer code, the response is calculated not as the evaluation of a regression function but, usually, as a series of algorithmic computations. Error or variation in the calculated response can be attributed to three causes. First, the algorithm itself might not be valid in the sense that it will not produce correct answers even with accurate input data. That is to say that the computational structure might be wrong. On this point, one would probably want to say a model is invalid if the discrepancy between observed and calculated responses is due to the computational structure of the code. Secondly, the values chosen for nonphysical inputs, like convergence criteria, time steps, etc., might lead to incorrect answers. In this case one could assume that best values are used for nonphysical inputs. Thirdly, the values used for physical and pseudophysical inputs could be wrong. Here a new question arises: Is a model "valid" if there exists a set of values (realistic or not) of the physical inputs such that the calculated response agrees with the observed response?

It is our opinion that a test of model validity should yield more than a yes or no answer. It should also attempt to pinpoint the cause of invalidity if the model fails the test.

MODEL APPLICATION

Statistical and probabilistic techniques can be used in the application of the verified/validated model. Steck (1975) covers examples in this area.

MODEL VERIFICATION

The area where we are concentrating our efforts extends into both the verification stage and the analysis stage. The problem is to locate troubles in the code and isolate the causes. In general, series of diagnostic or structural analyses (computer runs) are made with selected values of the inputs. If errors or inconsistencies are encountered, the physicist examines the results with the hope of identifying the cause. We are looking at this problem from the viewpoint of sensitivity analysis. The objective is to aid the physicist in his study by pointing out "important" inputs. As the number of inputs increases (it can easily pass 20) we feel that statistical techniques should take on an important role. A clarification of the definition of sensitivity and examples of some of our work are given later.

SOME CONSIDERATIONS

Before the statistician studies any of these problem areas, I think he should consider the following four points. First of all, the response variable may be a (parametric) function of time. If this is the case, statements related to verification, validation, analysis and application will have to be made both at specific time points and over the entire time history in general. In our work we have found codes that work very well in some time intervals and not so well in others. Secondly, the response is likely to be a very complicated nonlinear function of the inputs: experimental data is used to estimate parameters (inputs) in fitted functional relationships which subsequently appear in simultaneous partial differential equations which are solved numerically to obtain the response. Thirdly, when the code is applied to a "real" problem, the nuclear engineers may not know the precise values of inputs to use. Instead they can specify a probability distribution on the inputs which will change as better information is obtained. The final point is that the codes may be very costly to run.

A STATISTICAL MODEL AND OBJECTIVES

At this time I want to abstract the problem slightly and rephrase the statements of the areas of interest. Let $Y(t)$ denote the response (output) variable as a function of time, with $Y(t) = h(x;t)$ the computed values. Let X denote the vector of inputs to the code, and assume that they have a joint probability density function $f(x)$ for $x \in S$, the set of allowed values. The points the statistician might want to address are as follows:

(1) How well do the calculated values Y agree with experimental data?

(2) What is the probabilility distribution of Y? What is a (95%) confidence interval on the mean of Y? What is an upper, one-sided (5%) tolerance interval on the distribution?

(3) What is the sensitivity of Y to X? That is, how is a change in the value of X reflected in the value of Y and in the (conditional) probability distribution of Y?

The first of these three points has been discussed earlier and, although I think it is interesting and important, I will not address it again directly. Continuing then, one might ask if it is possible, or practical, to attempt

to answer both (2) and (3) simultaneously, in the sense that the same set of computer runs can be used to obtain information relevant to both points. With this question in mind, let us briefly look at three classes of techniques that might be employed.

CLASSES OF TECHNIQUES

PROPAGATION

In propagation techniques the moments of the distribution of the inputs are assumed given so that the moments of the distribution of the response can be obtained. (For example, in its simplest form linear propagation of error assumes that the moments of response can be adequately approximated by a first order Taylor series expansion about the mean of the inputs.) Derivatives are most often obtained numerically. The estimated moments of the response can be used to fit a distribution, and the derivatives can be used as an indication of sensitivity. An advantage of the procedure is that it requires assumptions on only the moments of the distribution of the inputs. The drawbacks occur in the complete process of approximating the distribution of the response variable.

ANALYTICAL APPROXIMATION

In analytical approximation techniques, a sufficient number of computer runs is made to enable a (nonlinear) response surface to be fitted to the output. Then, all analyses are performed on the fitted surface. The principal advantage of this procedure is that it allows for extensive investigations because of ease of evaluation of the new response variable. This is particularly important when one wants to study the effects of changes in distributional assumptions on the inputs. Drawbacks arise when good fits cannot be obtained.

SIMULATION

In simulation techniques the (transformed) probability distribution of the inputs is sampled to select values to use in computer runs. The resulting values of the response variable can be used to study the probability questions using standard statistical techniques. In addition, quantities related to sensitivity can be calculated. (Examples are given later). An

advantage of simulation techniques is that no specific functional relationship relating inputs and outputs need be assumed. A drawback is the possibility that an excessive number of runs may be required to obtain a specified degree of precision.

STRATIFIED SAMPLING

I am not discounting propagation techniques but, for what follows, I would like to focus on analytical approximation and simulation. In both techniques one of the first questions one might ask is how to choose the values of the inputs for the computer runs. In analytical approximation it is necessary to get a good picture of the response variable over the ranges of the inputs. Hence one might select a number of points within the range of each input via a classical experimental design such as a fractional factorial. In simulation techniques one should be aware that the range of each input might not be covered sufficiently to allow sensitivity analysis.

In our work in simulation techniques, we are using stratified sampling. The sample space S (the set of allowed values for the inputs) is partitioned into disjoint subsets which are then sampled according to some well defined rule. The reasons we are using this procedure are twofold. First, if the subsets are appropriately chosen, an adequate picture of the response can be obtained over the entire range of the input space with no regions "missed." Secondly, questions relating to the probability distribution of the response variable can also be answered.

To introduce the type of sampling schemes we are using, two examples will be given. As a first example, consider a case when the response $Y(t)$ is treated as a function of a single input X, with associated density function $f(x)$ for $x_L < x < x_U$. Assume that n simulation runs of the computer code are to be made. Define constants

$$x_L = a_0 < a_1 < \ldots < a_n = x_U$$

such that

$$\Pr\{X \in (a_{i-1}, a_i)\} = \frac{1}{n} \qquad i = 1, 2, \ldots, n .$$

The intervals (a_{i-1}, a_i) are called equal probability intervals for X. The n values $Y_i(t)$ of the response variable are obtained by sampling once from each of the intervals (a_{i-1}, a_i). Note that Y_i is a "random" Y for X on interval i.

For the second example consider a case where the response Y(t) is a function of <u>two</u> inputs X_1, X_2. Let the range of each X_i be divided into a lower interval L_i and an upper interval H_i where

$$Pr\{X_i \in L_i\} = Pr\{X_i \in H_i\} = \frac{1}{2} .$$

The four "random" values of Y are obtained by sampling (X_1, X_2) from the four subsets $[L_1, L_2]$, $[L_1, H_2]$, $[H_1, L_2]$, $[H_1, H_2]$.

A generalization is clear: Take a fractional factorial design and instead of using fixed values for the factor levels, use intervals for sampling purposes. Following this procedure for K observations yields K values of the response at K distinct values of each input. In addition, although the design is not completely symmetric in the input space, it should still allow for reasonably good fitting of surfaces for analytical approximation techniques.

I conclude this section on stratified sampling with a few questions. ("Level" refers to an interval which is sampled <u>each</u> time a value comes from it.) Since we are free to choose the <u>number</u> of levels for each input variable, how should this be done? What specific fractional factorials should be used? Given the total number of simulation runs that are allowed, are there optimal designs? (Some work has been done in this area by Conover, 1975.) What about "random" factorial designs (see Zacks, 1963, 1964, and Ehrenfeld and Zacks, 1967)? Can anything be gained by using other than equal probability intervals, as in importance sampling?

EXAMPLES OF SENSITIVITY ANALYSIS

At his point I would like to show some of the results from the application of statistical techniques to a computer code developed at the Los Alamos Scientific Laboratory by Hirt (1975) in the Hydrodynamics Group.

The problem is one of sensitivity analysis with the objective being to see if statistical techniques can be used to aid the physicist in his investigations and evaluations of the code. We carried out analyses of several response variables recorded at 0.1 millisecond (ms) intervals from 0 ms to 500 ms. We had four input variables and used a complete 2^4 factorial design. The low and high levels of each input were sampled using a uniform distribution. The statistics computed were the mean and (ordinary) standard deviation for the response variables, and, to measure sensitivity, partial rank correlation coefficients. (At each time point the 16 values of the response variables and the 16 values of each input variable were transformed to 1, 2, 3, ..., 16.) Using only these statistics we were able to identify the times at which the code was "active" and which of the inputs predominate as a function of time.

ADDITIONAL RESEARCH AREAS

In closing I would like to suggest a few more areas where research is needed: (1) General measures of sensitivity and ways of comparing them; (2) Sequential sampling plans which "learn" from previous simulation runs; (3) Methods for identifying important segments of input ranges; (4) Application of multivariate analysis techniques.

REFERENCES

Conover, W. J., Private communication, 1975.

Ehrenfeld, S., and Zacks, S., "Testing Hypotheses in Randomized Factorial Experiments," Annals of Mathematical Statistics, 38 (1967), 1494-1507.

Hirt, C. W., and Romero, N. C., Application of a Drift-Flux Model to Flashing in Straight Pipes, LA-6005-MS, Los Alamos Scientific Laboratory, 1975.

Kleijnen, J., Statistical Techniques in Simulation, Marcel Dekker, Inc., 1974.

Mihram, G. A., Simulation, Academic Press, 1972.

Steck, G. P., "How Should a Loss of Coolant Accident by Studied," Proceedings of the ERDA Statistical Symposium held at Los Alamos Scientific Laboratory, Los Alamos, New Mexico, November 3-5, 1975.

Tomovic, R., and Vukobratovic, M., General Sensitivity Theory, American Elsevier, 1972.

Zacks, S., "Complete Class of Linear Unbiased Estimators for Randomised Factorial Experiments," Annals of Mathematical Statistics, 34 (1963), 769-779.

Zacks, S., "Generalized Least Squares Estimators for Randomised Fractional Replication Design," Annals of Mathematical Statistics, 35 (1964) 696-704.

DISCUSSION: LOSS OF COOLANT

Conover: I have a question for George Steck. Your estimated function $\hat{Y}$ is remarkably similar to the unknown function Y. This indicates to me either great insight or tremendous luck. Could you describe the process you used to arrive at this function $\hat{Y}$? Also, could this method be converted into an algorithm that can be put on the computer?

Steck: We ran 45 observations and three of them were at 3,500. That made us think right away that since the largest one was 3,500 and we got three of those, we would pick 3,500 as the largest and try to fit the rest. So we threw away those and did our regression analysis on the remaining 42.

Now as for as the C-15 goes, in the beginning we had a series of levels of the factor C that fortunately straddled 15. In the second fractional factorial we narrowed those ranges and still straddled 15. When we plotted the functional values as a function of C alone, it looked like it might blow up somewhere around 15. I don't think we bothered to look at 14.9 or 15.1 to verify that; but we came in a little closer and it still looked like 15, so we just put down the C-15 in our regression model. It fit well enough so we left it.

We found the 3,500 just because that's the largest one that we observed; we found the C-15 by plotting techniques. As far as generalizing it, we did have an interactive computer graphics program written so that we could plot the functional values or the residuals against any kind of FORTRAN stateable function of the input, including the reciprocal of C-15 if we desired. But it never seemed to be quite as useful in the second problem as it was in the first one because we didn't have a pole in the second problem.

Conover: How did you pick your different values for your parameters, a, b, c, d?

Steck: At the beginning, we had an input range for each one. Assuming a uniform distribution over that range, we picked a 5% and a 95% point as lattice points in the fractional factorial. We do that for each of the six variables. Then the main effect regression shows that certain ones are more important than others. In a second factorial we brought the limits in by a factor of 3/4 or 1/2 around the nominal value. Why pick 3/4 or 1/2 or whatever it was? I am not sure; it just seemed a reasonable thing to try.

What is the effect on the regression procedure of the adaptive factor levels that we employed? We looked at the main effects, which suggested what the more important parameters were. Then we would try to complete a larger fraction on those important ones. Perhaps some interactions were aliased or looked like they might be important. So, we picked a few more points to try to get rid of the aliasing, or confounding, or whatever it was that we were being plagued with. We never took very many observations at a time. It was a sequential procedure, maybe a couple more here and a couple more there. Then we completed this fraction and completed that one until Warren Lyon came to visit us and we stopped.

Lyon: Steck didn't bring out one portion of his surface fittings that I think should be mentioned. I took a look at the behavior of his fit to the actual function in the vicinity of that pole. I threw out the 3,500 and just looked at the overall behavior. The fit was phenomenal! Even close to 15 where the function was in the million range the fit only missed by something like 0.7%. I was just astonished. What hasn't been mentioned is there was a second pole in the function that was missed. We took a quick look there as well and the behavior was quite different. It's easy to understand how it was missed. In the case of the pole with the variable at 15, the function behaved in a rather nice asymptotic manner and took off smoothly. In the case of the other pole, where one of the other variables, I believe, was equal to ten, there was an extremely sharp spike. The dependent variable was almost identical on both sides of the spike, so it was easy to see how they missed that. The overall implications probably were negligible, but I thought there was an extremely satisfactory fit on the first try. On the second function, which more closely approximated the behavior we will be investigating, again the fit seems to be very promising with an error on the order of a few hundred degrees. We are going to have to do much better, but I think overall the approach does look quite good.

Conover: I would like to comment on Mike McKay's presentation. I think Mike minimized his use of the partial rank correlation coefficient. He didn't explain it and I don't think its well understood since this is the first time I have seen anybody use it. He merely followed the same procedure that one would follow with regular partial correlation. You've got the Pearson product-moment correlation coefficient. You can form your matrices of correlation coefficients and get your partial correlation coefficients, as explained in many textbooks. Rank correlation follows the same procedure. The reason he used rank correlation, I think, is because he wasn't sure about some of the distributional properties of the input. He correlated inputs with outputs using rank correlation. Some of the inputs may be correlated with other inputs. Consequently, he may pick up an extraneous or superfluous correlation just because it is built-in; i.e., where X_2 may not be related to the output at all but X_2 happened to be correlated with X_1. The rank correlation coefficient then was put into the same formulas that you usually have for partial correlation to get this partial rank correlation.

One of the striking features I noticed in the output was that the rank correlation for one particular input was rather high and then it would drop off. Another rank correlation was rather low and it might go up to a certain point, which suggests to me a procedure for choosing your inputs going into the system. Steck used a sequential method for choosing inputs. I think that it is probably optimal to go with some kind of a sequential method; but, if you don't have the opportunity to use a sequential method or if you are forced to rely on intuition, which many of us are forced to do, it may not be quite as sharp as George's. You may want to choose your inputs in such a way to follow some sort of a pattern ahead of time.

Consider a one variable situation where you have Y as a function of X. In these computer codes, you can assume a monotonic function. This simplifies things considerably, particularly regarding the division by zero situations

that may arise. Suppose you have a monotonic function and the question is how big should X be in order to get the best Y value estimate as a Y function. The range for the X values may be given as a piece of information that you know. Rather than take a random sample of X values, it seems that to get the best information regarding the Y value, you'd want to make sure that you were taking a value of X from each region as you go along. Then you can observe some Y values and get a pretty good range on your function. In a one variable situation it seems intuitive that you would want to take a complete range of input values to get the most information regarding your function.

Now when you have two variables, the question comes up, "How can you do this?" Well, if you have a second variable involved here, you can follow the same procedure. But how do you pair X_1 and X_2 together? You have so many different combinations and you don't have that many runs available. This would suggest some kind of a random pairing, randomly pairing one value of X_1 with one value of X_2 and running them through. Repeat this so that you end up with say ten pairs. The result is you have sampled the entire range for X_1 and for X_2. If it turns out that X_1 is a variable that is important (such as may show up in your correlogram), then you have sampled the entire range for X_1. Later on if X_2 is the important variable, you've got the entire range for X_2. I just throw this out as a suggestion for a method for picking input variables to go into the system.

Friedman: Relative to techniques of statistical analysis of the output from large computer codes, you essentially have a problem of multivariate interpolation; namely, you extract a series of responses from your computer codes from various inputs. These of course are very expensive. What you would like to do once you have these responses is interpolate between them to understand the nature of your function over a full multivariate space. These techniques are referred to in the engineering literature as repromodeling, mainly pioneered by Bill Misell. He approaches this exact problem, has an algorithm for doing continuous piece-wise linear approximation to arbitrary multivariate functions, and has achieved dramatic successes.

Beggs: The problem is how to investigate loss of coolant accidents. Coolant loss causes a problem with temperature. We ought to address ourselves to find out how high the temperature will go. Two problems were presented: McKay deals with how to verify the model and sticks with the extremes. How well can you reproduce what happens in practice? McKay talked about stratified sampling and factorial design applications, using input variables to try to find out how the model itself is verified. You are concerned about how well you can predict maximum temperature. The only way you can do that is to try to search in the X space to find those X values which will give you the high-temperature readings. To do that efficiently, since we only have so much data, you use a steepest ascent or some kind of optimization search routine to try to find that region of X's that will give you the high temperatures. You can then look at the tail of the temperature distribution.

Easterling: I just want to emphasize what George (Steck) said about changing levels on our factorials. In the literature, we are used to seeing sequential

fractional factorials. There are certain ways in which you can build them up, i.e., to find sequences which do things in the ways you want them to. But in doing sequential factorials, we have to take the same levels. So if we have two variables, each at two levels, the standard way to run a sequential factorial would be to run the two corners of a square and then follow it with the other two corners. We changed these levels. The concept is that we might run two corner points and then come in and run two other corners of an interior square. Now the question of interest is when is this a good thing to do and when is it a bad thing to do. For example, suppose the effect of one variable, X_2, is not important. With this design, if we change levels, we now have a design which can pick up curvature with respect to X_1, whereas, if we had run the standard sequence we could not pick up curvature. I think this is an area for research - to consider sequential factorials with changing factor levels. Under what conditions is this a good idea, and under what conditons is this a bad idea?

Conover: There is one comment I have to make on comparison of the two methods. In one method, you look at a set of points, see what values you get for your function, and then take another set of points and so on. This is a superior method for finding the best function to fit the points. I don't see how you can use it to make probability statements concerning the mean. To make some confidence interval statements or other probability statements, you almost have to go to a preset design where you have the design planned ahead and there is some random sampling involved - sampling in such a way that every possible input has some known probability of being selected.

Easterling: We are not even thinking about the probabilistic properties of the output or how to evaluate functions, but rather how best to explore the X space to find this approximate function.

Tukey: Antithetic sampling is a regression sampling game that's due to John Hammersley. One approach might be to break your interval up into subintervals. Take two within each of these subintervals, one at random and the other symmetrically. This is the simple version and we get a certain amount of balance and also randomness.

Let me comment on a couple of minor points in the two papers and then I will make a general comment. We will see how much furor we can stir up. George Steck states, "This time though from 7,000 runs on the true response we could only simulate 42 sets of 165 for Monte Carlo comparison ..." One was really interested in average values here. He talks about quartiles of the 95% point estimated, but really those are average values too. I just think, George, you were a little bit timid. I would have been tempted to draw a lot more than 42 sets of 165 from 7,000 sampling with replacement. I don't see why I have to sample without replacement when I've got a population of 7,000 to play with. Now variability questions are a little bit touchy. I think there are ways to handle it. But it seems to me, one ought not to tie oneself down completely to independent repetitions.

McKay asked in his paper, "Can anything be gained in using other than equal probability intervals?" My main feeling about these three papers taken together is that people have not quite focused their ideas as to what they really wanted to find out and whether they were willing to pay what it took to find that out. If we believe things are monotone, roughly middle-humped or something, we certainly ought to concentrate our sampling accordingly. There's probably one end we don't care much about, except to know if it is low or high. We can't concentrate toward the other end and some in the middle just to be sure. I'm not arguing for concentrating everything. There is some merit to giving probability almost everywhere. But that merit is just a little bit limited. This sort of problem goes through a very complete change, somewhere between two and five dimensions. If you are going to do a problem with more than five parameters, the space is just not like anything you are used to thinking of. You will probably be stuck no matter what you do and if you talk about 50 input variables, your only hope is that four of them, you know not which, will be important. Then maybe you can get out and find something. Otherwise you will have to have a fairly large number of samples.

I would like to suggest that you ought to have a full code and a short code and a fitted function. The short code is something hopefully you get made up by the man who wrote the full code. Tell him he isn't supposed to get the right answers but he's only supposed to have 3% of the running time and to do as well as he can. Then on a sampling basis, you are in a position to look at the difference between the full code and the short code. You can run the short code well enough to do a reasonable job of fitting the function and you can run the fitted function to death if need be.

Our real problem is how to assess things out of a combination of values. For the fitted function, you can know what the distribution is as well as you might want to. Nobody is ever going to know any one hundredth of 1% points. You need to know how to apply the differences between the short code and the fitted function to say how is this going to be changed. At that point, you will probably have to go at least as far as to look at a plot of differences against fitted function values and be prepared for the fact that they show a considerable trend and some kind of spread around this trend as, for example, in Figure A. Here, you have lots of points and it isn't too difficult a job. But this smoothing technology, mentioned briefly yesterday, is going to be important to pick up things like this wiggly trend.

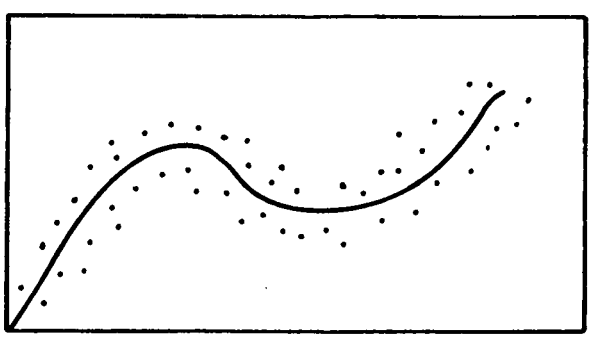


FIGURE A

We saw this morning, in a different kind of context, what happens to people who fit polynomials. Their necks usually end up on the block. A good smoothing technology will keep your neck from being chopped off. I don't believe there is any good polynomial technology. One can take whatever you fitted out here, look at the residuals, get some idea as to whether the spread of the residuals is reasonably constant, and try to use that part of the information to transform this distribution into what you think the short code distribution ought to be. You have some chance to check that one and see if your technology is working.

At the question of dealing with the differences between the full code and the short code, life is worse. But I don't see anything better to do than what was just discussed. If you can't run the full code a good many times, the only thing you can do is to bootstrap somehow. It seems to me this is the sensible sort of bootstrapping. If the short code is any good maybe you can run the short code enough so that then you can use the answers of the short code to determine the probability that you will run the full code, and then be in a position of weighing things accordingly. To me that is quite playable, certainly at the fitted function-short code interface.

The two points of view I would stress strongly are: (1) Let's put in another step or two, if we can get somebody to produce something reasonable for us at that point; and (2) Let's try to build up and look for things a little bit wilder than we thought they were, and try to use each of the more manageable things to guide what we do with the next less manageable things.

Steck: That's a good idea. In this problem it might be possible to get a short code because part of the long running time is the fact that you have numerical solutions to differential equations and small time steps. You can take a coarser time step and get something that is less realistic but shorter.

Tukey: In the case of one of those examples where there were data available from 0 to 500 milliseconds on .1 millisecond intervals, I think a logarithmic time step would probably work. All the fast action was in the first few milliseconds. After that, things got more and more systematic. You would shock people a little if you integrated the differential equations with respect to log t. But you might get a pretty good short code if you went that route.

Steck: I would like to return to Conover's comment. Somebody I knew made up the function, and I know something about him. He's in a hurry, so when I see 15, maybe I lock in on 15. He said afterwards that if you were going to do the problem again you might choose a random uniform number to put after 15, so it will be 15.4753 or something. But he had picked an integer and in checking the fit, I tried 15 and 15.1 and 14.9 in the fit. Fifteen was better than either of the other two and so I

picked 15; but if I had known in advance that it was a random process to produce this function, I wouldn't have settled on 15. I would have tried to search in the neighborhood to find the best one. But since a human had done it, I settled on 15.

Lyon: I have heard several excellent suggestions, some of which we are already trying or plan to do. First, what do we want? To establish that, I need to define two items. We have two basic computer programs. One we call best estimate, which is an attempt to predict as accurately as possible what happens. The second we call an evaluation model program, which is very similar but which contains a number of conservatisms which make it over-predict temperature. We use this to evaluate license applications and to evaluate analyses and compare against the regulatory requirements. What we want first is a quantification of the best estimate program. Secondly, we want to be able to compare the best estimate and evaluation model programs in some sort of a quantified manner. Thirdly, we want to be able to make quantitative statements in the application of these programs to accident investigations. I think we have an idea of what it is we want. I am not sure we know how to get there.

On the comments of the short code, I am not sure we are bright enough to write a computer program that is shorter than the monstrosity that we presently have with the knowledge that we have available to us and still put in the incredibly complex phenomena that are taking place. That does not say that we cannot have a shorter version, because there are a couple of things available to us. One possibility is the time step. There is another possibility. Let's call it a mesh size or a nodalization, the number of items you choose to represent the system. In the time step, one of the problems is that in order to maintain numerical stability you are stuck with a time step. We do indeed use a very fine time step in many portions of the investigation and we use a gross time step when we can get away with it. Fortunately, when you simplify the nodalization in general you can increase the time step. Of course accuracy suffers, principally because of the nodalization selection. The next step in our investigation, already under way, is the use of many fewer nodes, so that the program runs much more rapidly. In an overall investigation we are still talking in the order of 1/2 hour to an hour per point, so while we may be saving a lot of time it's still quite expensive. (Post Conference Comment: We are modifying the computer program so that different time steps can be used for different nodes. This should improve running time but not to the point that the problem is eliminated.)

One advantage of the Steck and Easterling approach to response surface generation is the use of many different types of functions. They have not limited their approach to X and X^2, but also use logs and exponentials.

We also had a comment on a desirability to investigate all variables. I wish I knew how to do this. Literally there are thousands of independent variables in this problem. We have variables associated with the input,

such as what times are we going to investigate during the history when events are occurring? At what power level are we operating? What is the time power history? What are the dimensions of many of the items within the system we are simulating? We have dozens of correlations built into this program. We have uncertainties associated with these correlations, both in their fit to the original experimental data and in their fit when they are used outside of ranges for which they originally were derived. I don't know how the statistician can help us with that one. The approach we probably will follow is going to depend somewhat on an engineering feel for what is important and what is not, backed up by some investigations. And hopefully with a lot of help from people like Bob Easterling and George Steck.

Tukey: From what I heard of your comment, you don't seem to know what you want to know. Do you want to control the 50% point, the 90% point, or the 95% point? You are going to want to take the data differently depending on those. It just isn't feasible in a messy problem to take the data so as to get high grade information about a lot of different answers. The answers have to be more refined. Regretable but true!

Let me just remark for the benefit of the statisticians present, that the word correlation in your comment was, I think, used in the engineer's sense and not the statistician's sense. That maybe needs to be cleared up. I think for something like a short code to be helpful, you need not get the right answer. That's the difficult point to get across to the man who's writing the code. It's supposed to get answers that are well correlated with the right answer in the statistician sense. But it doesn't have to be right. That means you can probably do a lot more violence to it. You would be surprised how much violence I'd be willing to try and see how it did.

Uppuluri: At times, one needs to resort to Monte Carlo methods to study distributional properties of functions $f(X_1, X_2, \ldots X_n)$ of n random variables. In general, one has information about the moments of $X_1, X_2, \ldots X_n$. It may be a good practice to compute the moments of f and compare with those based on the Monte Carlo.

Relative to remarks about using no more than three to five parameters to make the whole scheme depend on them, I would like to know whether this is based on the catastrophe theory proposed by René Thom.

Tukey: I am afraid I don't know the references I am asked to comment on. The sort of trouble I am talking about is that large dimensional spaces are larger than you think they are. Trying to do things regularly gets you in trouble because of the number of points that you have to use. It's a nasty business. Let me also comment that I take it Monte Carlo here today means what I call experimental sampling rather than more sophisticated Monte Carlo.

Lyon: I agree with your comment regarding messing up the code by highly simplifying it. However, to be useful we still have to have some semblance to reality and we are already stretching it. The work going into the code itself plus the changes might be very expensive. So far, we are just following the route of reducing the number of nodes to describe the problem, until we start to deviate so far from reality that we figure we'd better stop. If we limit ourselves to only four or five independent variables, I personally feel that we cannot accomplish the things that we need to do. We have to have an approach that lets us look in the realm of 20 to 50 variables. That is a tough assignment. I am not sure we really know how to do that yet.

Tukey: I think one point is perhaps slipping by and let me make myself clear on another one. If you have a short code and you use it with the full code, you are not relying on the short code any more than you are the fitted function. You are relying on the full code and you are trying to get every darn ounce you can out of each one of the full code points. It seems to me that this is what we are trying to do. I would be willing to cut those nodes down below physical reality and see whether it helped me enough.

On the number of variables, I was not arguing for putting few variables in. What I was saying was that unless you are lucky enough after you put 50 variables in to find out that too many of them make big effects at the back end your neck is probably on the block anyway. By all means put the 50 in and let's hope that four of them do most of the trick. Then the rest of the 46 can be sort of picked up as a blur. That's what you have to do. I don't think we have any real difference of opinion here. But I do think that if there are 12 equally important variables, the chances that you get a good answer with a small number of runs of the full code are not very good.

Lyon: Thank you. Your first suggestion is an excellent one. I think we will probably try that during our investigations. On the second one, I have an engineering feel that we are going to be stuck with a number of variables which are quite important. We'll find out.

Steck: The basic output of the code is not just a peak temperature but rather a time history. I showed a little drawing of a two humped thing, like a camel without a head. One thought that I had was to fit each transient, each double hump thing, with maybe 11 nodes in splines. We actually carried that out very successfully. You express the transient then as a sum of splines with coefficients that are determined and the polynomials are fixed. You can imagine doing a regression on the coefficients of the spline. At the end you will be able to sample from the inputs and get a function out rather than just a peak temperature. Then you can maybe formally differentiate the spline and find out where its maximums are and pick the largest one of those. Has anybody had experience with tackling any problem like that where you fit with splines and then do a regression on the spline coefficient?

Tukey: Were you regressing something else on the spline coefficients or were you regressing the spline coefficients on the input data?

Steck: Regressing the spline coefficients on the input data.

Tukey: I don't have experience with splines in this, partly because I am pessimistic about them. I don't like a fitting procedure which will change things over here at the maximum in order to make a better fit over here at the minimum or over beyond the minimum somewhere. That's what the continuity constraints on the splines lead you to do. It doesn't seem to me they are really localized enough to make me happy. I foresee some difficulties, but maybe you will come out all right.

Steck: I think in this case I come out all right. I have maybe a dozen plots of the transient produced by the full code and comparable results obtained from regressing the spline coefficients on the input data. You are hard put to see any difference between them. It fits very well. The thing you are fitting is smooth enough and I guess I picked enough nodes so that there wasn't any problem.

Pool: I direct the mathematical and computer science research programs at the Argonne. I am here trying to find ways that statistics can interact with some of the things we are doing. For the last several years we have had a research program in the evaluation of numerical software. Not an evaluation of hydrodynamic codes or anything like that. We are much more pedestrian than that. We started out with elementary functions, special functions and worked our way up to numerical linear algebra where you have a backward error analysis to depend upon. We moved away from that case to where you do not have an error analysis. The first step is quadrature, where you have a user supplied function that the quadrature subroutine must use. So all the error analysis techniques go out the window, unless you know the general analytical form of the functions that the user is going to supply. You usually don't. We've started to look at the minimization software where you have the same kind of complications plus the fact that the algorithms interact themselves in a relatively sophisticated way. We started to look at ordinary differential equations software in conjunction with people from Livermore and Sandia.

I really wonder whether the step between things like optimization and minimization software to a hydrodynamics code isn't a bit of a giant step. I was wondering whether any of the people have tried to do something intermediate, where perhaps some of the numerical analysts error analysis capabilities would allow you to verify the kind of inferences you are making with statistical techniques. It makes me very nervous to hear an analysis of loss of coolant codes. For example, little is known about something as simple as the coupling of linear algebra codes. I know we've never done anything like that, although we know a great deal about individual linear algebra codes.

Tukey: On the forward integration differential equation codes, the backward error analysis to the initial condition is really what you want.

DIELECTRIC BREAKDOWN IN LIQUID HELIUM

F. L. Miller, Jr.
Union Carbide Corporation Nuclear Division
Oak Ridge, Tennessee

THE PHYSICAL SETTING

Reduction in the cost of underground electrical transmission is important if the utility industry is to meet the nation's needs for electric energy without undue cost while preserving the amenities in and near load centers and power plants. Cryogenic and superconducting cables are being developed and the problem which is discussed here arises from an attempt to understand engineering aspects of dielectric breakdown in liquid helium.

The experimental apparatus consists of a 130 KV(dc)/80 KV(ac) intermediate voltage unit and a 600 KV(dc)/700 KV(ac) high voltage unit under construction. The data discussed here come from the intermediate voltage unit. Leaving aside the necessary measuring equipment and electronics, the experimental devices consist of an insulated container, or dewar, in which two electrodes are placed, one above the other. The shapes of these electrodes can be altered to experiment with different magnetic fields and the distance between the electrodes can be varied to suit the experimenter. The dewar is then filled with liquid helium and after the equipment has reached temperature equilibrium the measurement process can begin. A voltage is built up in one electrode until an arc occurs to the other electrode. This is also called a flashover and the voltage attained when the arcing occurs is called the breakdown voltage. Conceptually, this process can be repeated an arbitrarily large number of times. Practically, there are limitations of time, personnel, buildup of impurities and damage to the electrodes.

Experiments like these have been performed many times on transformer oils. It was found that the first few measurements needed to be discarded to allow for breaking in the system and after a series of "good" measurements (commonly about 20) the breakdown voltages decreased as a result of impurity buildup and electrode damage. Much the same scenario occurs here. A few measurements are taken and discarded during a system checkout phase

and a degradation of later measurements has been noticed in rapidly boiling liquid helium. This has been stopped by carefully maintaining the helium level and by slightly pressurizing the dewar to inhibit bubble formation.

THE PROBLEM

Figure 1 shows a typical set of data. Our "standard" practice is to check for evidence of decrease in breakdown voltage with time and occasionally to run a set of data through a fast fourier transform to see if any periodicities appear. Then as in Figure 2, the data is plotted on several probability papers.

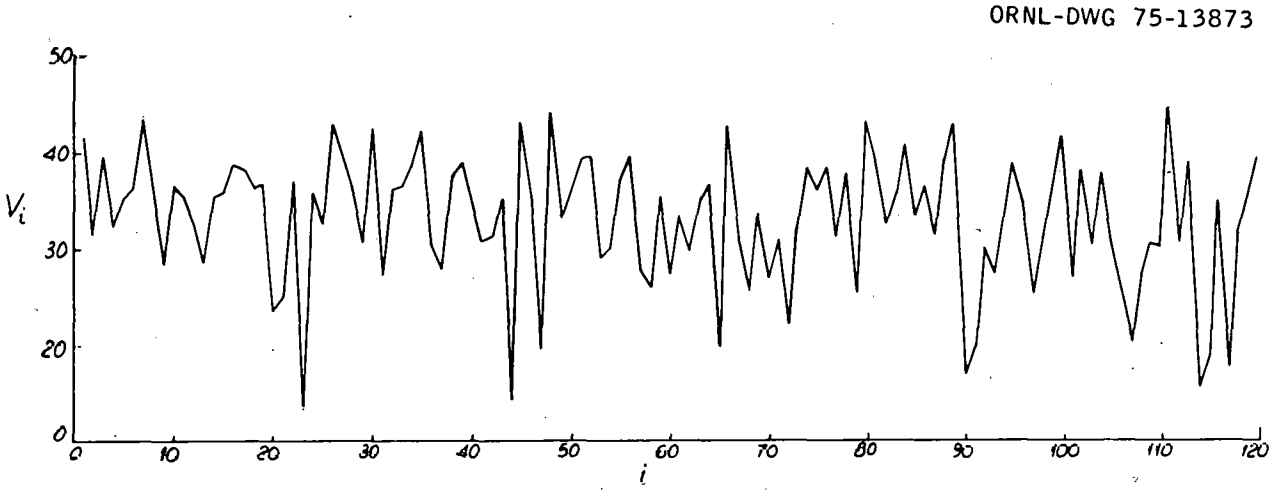


FIGURE 1. Data Taken at 0.5 mm Gap Width on March 11, 1974, with Sphere Positive, dc.

Several distributions for voltage breakdown have appeared in literature. They range from the normal (Henning and Wartman, 1957) and lognormal (Gerhold, 1972) to extreme value distributions like the Weibull (Occhini, 1971) and the extreme value (Weber and Endicott, 1956). The normal distribution is a poor candidate as the data are usually skewed to the left. A case can be made for using one of the asymptotic extreme value distributions on the grounds that the voltage at which the breakdown occurs is the minimum voltage necessary, at this particular instant in time, to force an arc across one of the large number of possible arc paths between the electrodes. We have found that the

plot on log extreme value probability paper seems adequately linear for most sets of data. This is a somewhat subjective observation, one that I am not entirely happy with, but at least it is a starting point.

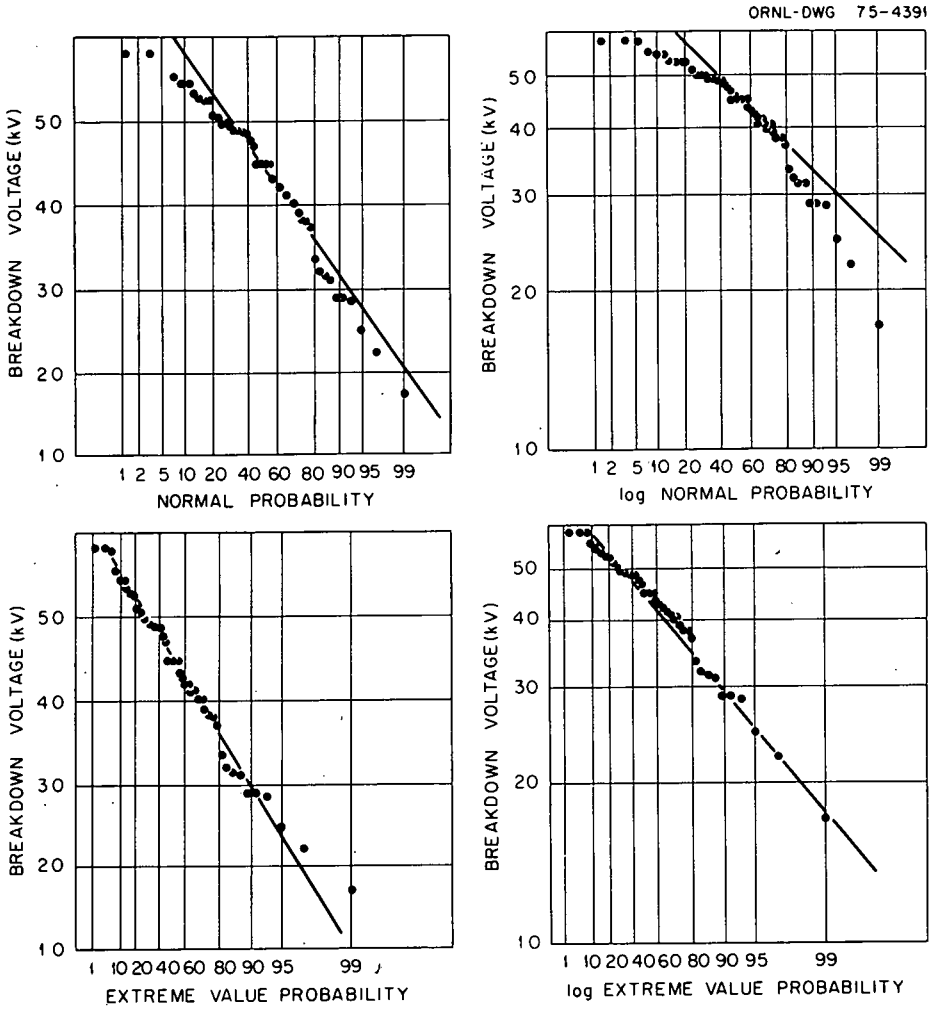


FIGURE 2. Probability Plots for 2 mm Gap, Point Positive, dc.

There are three asymptotic types of distributions for the smallest order statistics (Fisher and Tippett, 1928). Below are left-skewed forms of these distributions and some of their names.

Type 1, extreme value, Gumbel, log Weibull, double exponential:

$$F_1(x) = 1 - \exp\left[-e^{-\left(\frac{x-\xi}{\theta}\right)}\right] \qquad -\infty < x < \infty$$

$$\theta > 0$$

$$f_1(x) = \frac{1}{\theta} e^{-\left(\frac{x-\xi}{\theta}\right)} \exp\left[-e^{-\left(\frac{x-\xi}{\theta}\right)}\right] \qquad -\infty < \xi < \infty$$

Type 2, log extreme value, Rosin-Rammler:

$$F_2(x) = 1 - e^{-(\frac{x-\xi}{\theta})^{-k}} \qquad \xi \leq x < \infty$$

$$\theta > 0$$

$$f_2(x) = \frac{1}{\theta}\,(\frac{x-\xi}{\theta})^{-k-1}\, e^{-(\frac{x-\xi}{\theta})^{-k}} \qquad k > 0$$

Type 3, Weibull:

$$F_3(x) = 1 - e^{-(\frac{x-\xi}{\theta})^{k}} \qquad \xi \leq x < \infty$$

$$\theta > 0$$

$$f_3(x) = k(\frac{x-\xi}{\theta})^{k-1}\, e^{-(\frac{x-\xi}{\theta})^{k}} \qquad k > 0$$

The Type 1 distribution is a sort of transition from Type 2 to Type 3 distributions. Useful references for extreme value theory are Gumbel (1958), Epstein (1960) and Johnson and Kotz (1970).

The location parameter ξ is troublesome. It is dear to the hearts of the experimenters as it implies the existence of a voltage below which breakdown cannot occur. There seems to be no physical justification for such a parameter and the left-skewed observations provide evidence against its existence. Values for it have not yet been estimated for these reasons, but the idea is attractive. Figure 3 is a log extreme value probability plot of data taken during March 1974 at various gap widths. Reversing the polarity does have an effect on breakdown strength, contrary to physical theory. It is suspected that the placement of the electrodes one above the other is responsible for this effect. Another striking thing about this figure is the discontinuity near 90% in the data. The equipment was running well and there is no explanation for this phenomenon. The number of observations for each gap width and polarity are listed below:

Gap Width (MM)	Sphere +	Sphere -
.25		30
.5	120	180
1.0	60	
1.25	160	30
1.5	120	120
2.25	120	30

This apparent mixture of distributions does not occur in every set of data and no pattern for its occurrence has been detected. It is quite troublesome as engineering considerations for the design of cryogenic cables depend on probability statements about minimum breakdown strength.

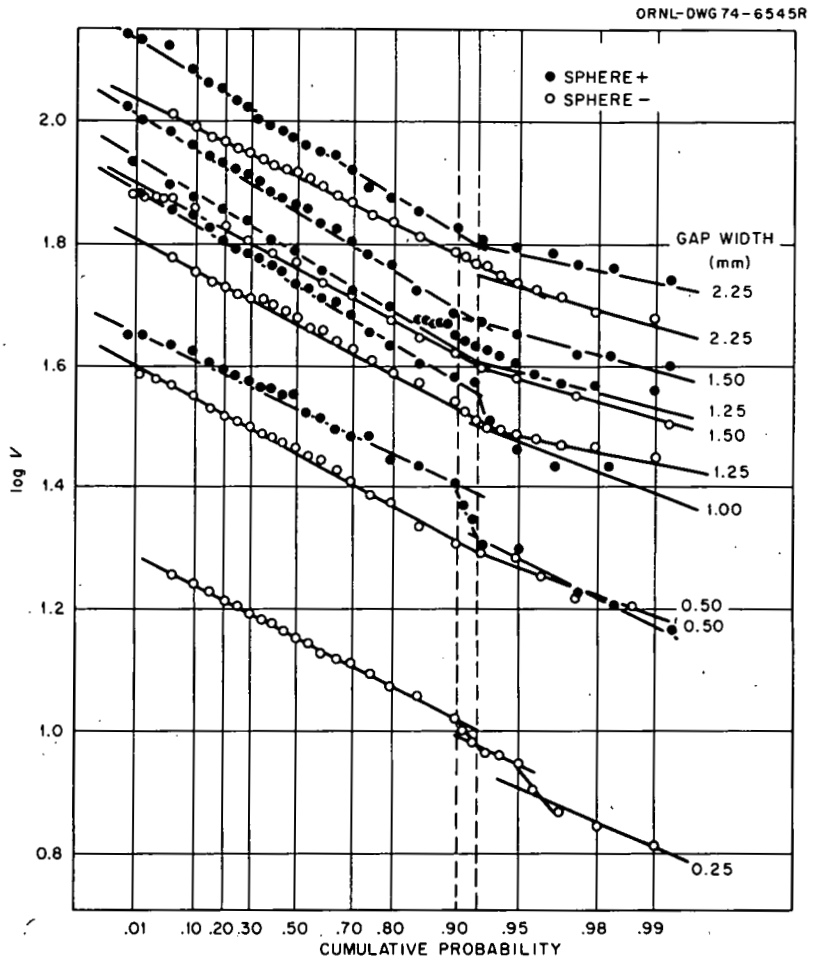


FIGURE 3. Log Extreme Value Probability Plots of Data Taken During March 1974, dc.

TOWARD A SOLUTION

The use of probability plots to determine the distributional form of voltage breakdowns cannot be considered definitive because of the nature of the method. However, when you consider the totality of the evidence it is more persuasive. What is needed is a test which distinguishes between the extreme value distributions. New experimental evidence must be found before the existence of a minimum breakdown voltage can be accepted. The factors which produce the contamination of the primary response need to be identified.

REFERENCES

Epstein, B., "Elements of Extreme Value Theory," Technometrics, 2 (1960), 27-41.

Fisher, R. A., and Tippett, L. H. C., "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," Proceedings of Cambridge Philosophical Society XXIV (1928), 180-190.

Gerhold, J., Thesis, Technical University Graz, 1972.

Gumbel, E. J., Statistics of Extremes, Columbia University Press, New York, 1958.

Henning, H. J., and Wartmann, R., "Stichproben Kleinen Umfanges in Wahrscheinlichkeitsnetz," Mitteilungsblatt Fur Mathematische Statistik, 9 (1957), 168-181.

Johnson, N. L., and Kotz, S., Continuous Univariate Distributions--Part 1, Houghton Mifflin, Boston, 1970.

Occhini, E., A Statistical Approach to the Discussion of the Dielectric Strength in Electric Cables," Transactions of the IEEE, Winter Power Meeting, 1971.

Weber, K. H., and Endicott, H. S., "Area Effect and its External Basis for the Electric Breakdown of Transformer Oil," AIEE, 75 (1956), 371-381.

DISCUSSION: DIELECTRIC BREAKDOWN IN LIQUID HELIUM

Editor's Note: The problem being considered concerns deciding between one of three extreme value distributions on the basis of probability plotting. Something is needed that is more satisfying and perhaps more repeatable than merely seeing if the set of dots are on a straight line. Are there techniques that would allow discrimination among these three distributions? Are there nonparametric techniques that would help in a situation like this?

Lohrding: Have you tried the goodness-of-fit, Kolmogorov-Smirnoff type test? There are probably more powerful tests on the market now, but it seems that that would be one of the first things to look at instead of looking to see if you have a straight line on the probability graph plot paper.

Miller: If you get into Kolmogorov-Smirnoff statistics, Lilliefors published papers on Kolmogorov-Smirnoff type tests in normal situations where the mean and variances are unknown and the exponential. I suspect that to create the necessary tables would be a heavy computing job. But it is one method that is open for pursuit. I'd like something a little better than Kolmogorov-Smirnoff tests because they are notoriously weak in the tails, the area of greatest concern.

Lohrding: There have been some recent tests that do pick up the tail difference much better than the Kolmogorov-Smirnoff.

Tukey: I don't think that a statistician should allow himself to be caught trying to decide between these extreme value distributions with data of the general character or even ten times better than that shown here. It isn't going to be settled that way. You can never find out about the far tails of things from small amounts of data. And in this business and for this far tail, 5,000 values is still a small amount of data. So my recommendation is do your best to avoid trying to solve that problem, because you aren't going to do it.

Bloomfield: I was looking at Figure 2. The sharp discontinuities and the small number of these graphs mask the gentle curvature that isn't in some of the others. These data are probably consistent with the observation that the logarithmic transforms have been a bit too powerful on the curves of higher voltages. The lower ones look as if they have been straightened out fairly well. The other ones look as if they have been turned over and are now convex, whereas before they were concave. Possibly a transform that looks like a logarithmic transform at the lower voltages and something closer to a cube root at the higher voltages would succeed in linearizing all these things at the same time. All three of these distributions are extreme value and essentially the same distribution. If you have a variable and take any one of those distributions and apply homogeneous power transforms to it or a logarithmic transform, then you get one of the others. This occurs because the extreme value problem that Fisher and

Tippett solved was phrased in the same way as the central limit theorem. But there is no real reason to stay within that particular class of transformations unless there is some natural invariance that you wish to preserve. The thing about those transformations is the homogeneity of them. If you make a scale change to one variable then it comes out as a scale change in the transformed variable, or shift in the special case of the logarithmic transform. In the present case, the data are collected on equipment which has some characteristic geometry and some characteristic scales. It is not necessarily reasonable to expect that sort of invariance to be a natural requirement. It would be quite reasonable to look for other types of transforms that may do a job on these data of simultaneously linearizing several of the plots.

Miller: The jumps do not appear in all the data that I have seen. This is "the best data" according to the physicists. Some of the earlier data and some of the later data do not show this discontinuity between 90 and 95% at all. I am really assuming that the physicists, with some help from concommitant observations perhaps, can find a way to eliminate this "contamination." I don't see any way to handle mixtures of distributions in a situation like this.

Tukey: I would rather look at better plots than the plots we are looking at here. I want to take out a linear fit from these things and look at the residuals. Then maybe I can tell better what seems to be going on. This is an engineering problem and the physical solution might not be the right answer. The properties of chemically pure copper are of no interest to the telephone business, because the copper in the telephone business is not chemically pure and never will be. The properties of liquid helium flash-over, to use the nasty word, that are going to be relevant to transmission lines are those that apply in the circumstances really in the transmission line. If the physicist can find out how to get rid of this jump here, the next question is, will that technique apply in the transmission line or will the transmission line have worse jumps for the same reason? I think you have to understand the phenomenon and not just clean it up. The phenomenon of Chile copper is still not yet understood.

Kao: Although I am in the Department of Applied Mathematics, I get involved with data analysis for probability distribution of pipe bursting. That is very similar to the problem of dielectric breakdown that we are now talking about. I will mention the procedure I suggested and have been doing for the problem that I have. First, I would search for the physical reason for using any kind of theoretical model like the extreme value distribution. There are three types of extreme value distributions; the three types are different. These three types of extreme value distribution are limiting distributions based on different assumptions. You have to check into the basic assumptions. I found out the Weibull distribution fit the basic assumption. Now the pipe burst and pipe bursting pressure are all positive, so it is bounded from below. If you check the smallest extreme value, it is bounded from below all right. I developed what I call densitygrams. The width of interval for plotting the step function is not fixed; it's determined by the sample values. Using that technique, I came up with a

curve close to a largest value extreme, not the smallest value extreme. That curve also looks like Weibull. For pipe bursting pressure there is no reason to take the largest value, so Weibull turned out to be the right choice. I used the Weibull and am now doing an estimate for the parameters. So far the results seem to turn out all right. It is almost impossible from an empirical point of view, as Professor Tukey just mentioned, to use just 17 points to evaluate tail probabilities. I don't really agree that we can never solve this problem. If there is a physical reason to justify picking some theoretical model, I think we still can try our best and find some results which may give us some insight in the preliminary stages.

Conover: One needs to decide what is the purpose of making a decision as to which probability distribution to choose. If you are looking for the best fit, then I would just use a likelihood ratio. Figure the likelihood function of one distribution, compare it with the likelihood function for another distribution for the points you have, and pick the distribution that comes up with the largest likelihood function.

If you are really interested in examining what is going on in the tails, an examination of your graphs indicates that none of these distributions is working well in the tails. Perhaps you should go with some kind of a polynomial function to fit the tails.

You mentioned the method used by Lilliefors. I don't think it would be too difficult to use it for your problem here. I am not sure that the end purpose is what you want. The end purpose will be merely a measure of goodness of fit. But if you want to go ahead with the procedure it is simple enough to do. Just estimate the parameters using some good method such as maximum likelihood. Generate samples of the same size from the distribution in question with those parameters. Do this repeatedly 5,000 times, however many you feel you can afford. This doesn't take up much computer time to come up with these numbers and see how many of these empirical distribution functions are further away from the theoretical function than your observed points are. I really don't think it will take much time. Lilliefors used a lot of time because he tried to generate complete tables that anybody can use. You are just interested in a particular set of values, a particular sample size. You only have to go through it once.

Kao: I didn't understand what Conover said about using the maximum likelihood ratio comparison for picking some probability distribution, could you elaborate?

Conover: I was suggesting fitting the unknown parameters using some reasonably good method that is not too difficult. Then compute the likelihood function. Put the observed values into the density function. Compute your likelihood function for the Gumbel distribution. Then do the same thing for the Weibull distribution. See which distribution gives you the larger likelihood function.

Tukey: I was not arguing against the use of physical insight and physical theory to try to do the extrapolations here. I would argue in favor of it. But I think the wise statistician will make sure that the physicist is the man who is doing it.

If you put polynomials into the kind of problem here, you are guaranteed either to get a nonmonotone curve for extreme extrapolations and somebody will extrapolate out there. Or else you will get an answer on the finite probability of failure at zero stress and somebody will extrapolate out there. So don't try the polynomial.

EXPERIMENTAL DESIGN FOR HTGR FUEL RODS

C. K. Bayne
Union Carbide Corporation Nuclear Division
Oak Ridge, Tennessee

PROBLEM

A screening experiment can be completed in three months in a High Flux Isotopic Reactor (HFIR) to test factors affecting particle coating breakage in fuel rods. This knowledge will be used to design fuel rods for the High Temperature Gas Reactor (HTGR).

HTGR fuel rods are formed into cylinders by injecting a pitch-graphite matrix over coated fuel particles. The fuel rods are then carbonized, with some of the pitch volatilizing. The amount of gas released is indirectly measured by the Pitch-Coke ratio (P-C ratio) of the weight of the carbonized matrix to the weight of the original matrix. Four matrices are examined which have P-C ratios ranging from 20% to 40%. Matrix A has a high P-C ratio and is expected to cause high coating breakage. Matrices B, C and D have low P-C ratios with matrix B being a standard matrix and matrices C and D alternatives to matrix A.

To test fuel coating breakage, two holes are available in a HFIR with each hole containing sixteen fuel rods (Figure 1). One hole is run for one cycle and the other hole is run for two cycles. The middle eight fuel rods in each hole contain enough fissionable material to create a temperature of 1250°C, while the four rods at each end contain enough fissionable material to create a temperature of 900°C. The amount of fissionable material in each fuel rod will depend on its position in the reactor.

An additional factor in the experiment is the flux ($n/cm^2 \cdot sec$) which varies parabolically over the length of each hole. The highest flux value is about 12×10^{14} MeV for the middle fuel rods and the lowest flux value is about 6×10^{14} MeV for the fuel rods at each end of the hole.

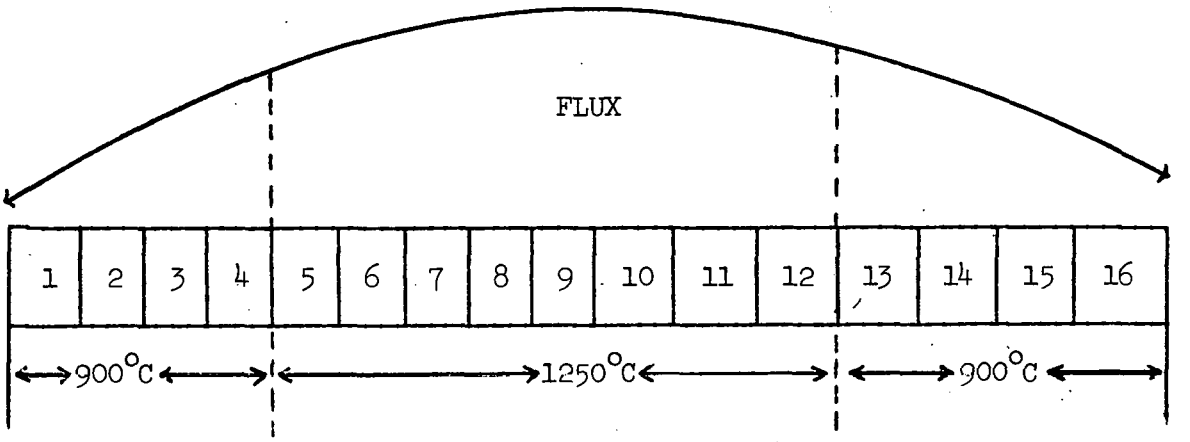


FIGURE 1. Horizontal Diagram of Sixteen Fuel Rods in One Hole

The experimental variables and their levels are given in Table 1.

TABLE 1. Levels of Experimental Variables

Variable	Level
Cycle	Cycle I (one cycle); Cycle II (two cycles)
Temperature	900°C, 1250°C
Particle Coating	High strength, Low strength
Matrix	A(High P-C ratio); B,C,D (Low P-C ratio)
Flux	Continuous

An experimental design is desired to analyze the effect of the variables on particle coating breakage. The solution is divided into three parts. First an experimental design is discussed for the factorial (discrete) variables: cycle, temperature, particle coating and matrix. Then the flux variable and the response variable are considered separately.

FACTORIAL VARIABLES

The experimental design for this problem must be constructed to satisfy several aims. Three aims which are considered most important are: (1) comparing the matrix A with high P-C ratio with the matrices B, C and D with low P-C ratios; (2) comparing the effects of the three low P-C ratio matrices; and

(3) testing the main effects and the two factor interactions of the four factors: particle coating (P), matrix (M), temperature (T) and cycle (C).

The effects of the factors with two levels are estimated as the difference between the response at the high and low levels. This means the parameterizations of factors C, T, and P are as shown in Table 2. The matrix factor is parameterized into three effects M1, M2, and M3 defined by the contrast of responses from fuel rods containing matrices A, B, C, and D given in Table 3.

TABLE 2. Parameterization of the Two-Level Factors

Variable	Parameterization
Cycle	+1 (cycle II), -1 (cycle I)
Temperature	+1 (1250°C), -1 (900°C)
Particle Coating	+1 (high strength), -1 (low strength)

TABLE 3. Response Contrasts of Matrix Factor

	Matrix			
Variable	A	B	C	D
M1	+1	-1	-1	-1
M2	0	-1	-1	+1
M3	0	-1	+1	0

The M1 contrast is a comparison of a matrix with high P-C ratio with matrices with a low P-C ratio. Since this comparison is of great importance, matrix A will be used in sixteen fuel rods while matrices B, C and D will occupy the remaining sixteen fuel rods. The second most important matrix contrast M2 is between a commercially made matrix D and matrices B and C. Therefore, matrix D will be assigned to eight fuel rods and matrices B and C each will be assigned to four fuel rods. The last contrast M3 is a comparison of matrix C with the standard matrix B.

The main effects are represented by the letters C, T, and P, and the contrast by M1, M2, and M3. The two-factor interactions of the main effects are represented by: CT, CP, CM1, CM2, CM3, TP, TM1, TM2, TM3, PM1, PM2, and PM3.

The experimental design for the factorial variables is constructed by first considering the matrix variable at two levels. The two levels are high P-C ratio and low P-C ratio. A 2^4 factorial design can then be constructed with two replicates. This design will indicate the level of the cycle, temperature, and particle coating variables for each fuel rod, and matrix A will be assigned to the sixteen fuel rods that are to receive the high level of the matrix variable. To complete the design, the low P-C ratio matrices B, C and D must be placed in the sixteen fuel rods assigned to the low-level of the matrix variable. To accomplish this, this part of the design is considered to be a 3×2^3 asymmetrical fractional factorial design in 16 runs. The low P-C ratio matrices are considered to be a three-level variable coded B = 0, C = 1, and D = 2. For construction purposes, the high and low-levels of the two-level factors are identified as 1 and 0, respectively.

To construct a 3×2^3 design in 16 runs, Addelman's (1962) replacement and collapsing method is used. Addelman's method utilizes the proportional frequency condition, $n_{ij} = n_{i\cdot}\, n_{\cdot j}/N$ for two factors X and Y, where N = number of treatment combinations, $n_{i\cdot}$ = number of times the i^{th} level of factor X occurs, $n_{\cdot j}$ = number of times the j^{th} level of factor Y occurs, and n_{ij} = number of times the i^{th} level of X occurs with the j^{th} level of Y. The proportional frequency condition was shown by Plackett (1946) to be a necessary and sufficient condition for two factors to be orthogonal to each other and to the mean (μ).

First a 2^6 design in 16 runs is constructed with the three variables: C (cycle), T (temperature), P (particle coating), and three dummy variables X, Y, and Z. The $2^6//16$ design is constructed using the identity relation: $I \equiv CPTX = CTYZ = PXYZ$ (Table 4). This relationship forms a resolution IV design which is the maximum possible resolution. In resolution IV designs, the mean and main effects are orthogonal but the two factor interactions are confounded.

TABLE 4. A $2^6//16$ Design with I = CTPX = CTYZ = PXYZ

C	T	P	X	Y	Z
1	1	1	1	1	1
1	1	1	1	0	0
1	1	0	0	1	1
1	1	0	0	0	0
1	0	1	0	0	1
1	0	1	0	1	0
1	0	0	1	0	1
1	0	0	1	1	0
0	1	1	0	0	1
0	1	1	0	1	0
0	1	0	1	0	1
0	1	0	1	1	0
0	0	1	1	1	1
0	0	1	1	0	0
0	0	0	0	1	1
0	0	0	0	0	0

The three two-level dummy variables are now replaced with a four-level variable S by the correspondence shown in Table 5. This correspondence is used because an orthogonal main effects plan for three two-level factors can be replaced by a single four-level factor and the proportional frequency condition is preserved (Table 6).

TABLE 5. The Correspondence used to Replace Three Two-Level Variables with One Four-Level Variable.

Two-Level Variables			≠	Four-Level Variable				
X	Y	Z		X	Y	Z		S
0	0	0	and	1	1	1	⋯⋯→	0
0	1	1	and	1	0	0	——→	1
1	0	1	and	0	1	0	⋯⋯→	2
1	1	0	and	0	0	1	⋯⋯→	3

TABLE 6. A 4 x 2^3 Design in 16 Runs

C	T	P	S
1	1	1	0
1	1	1	1
1	1	0	1
1	1	0	0
1	0	1	3
1	0	1	2
1	0	0	2
1	0	0	3
0	1	1	3
0	1	1	2
0	1	0	2
0	1	0	3
0	0	1	0
0	0	1	1
0	0	0	1
0	0	0	0

For example, the proportional frequency relationships for the factors, P and S are:

$$N_{10} = 2 = N_{1\cdot}N_{\cdot 0}/N_{\cdot\cdot} = 8 \cdot 4/16 = N_{0\cdot}N_{\cdot 0}/N_{\cdot\cdot} = 2 = N_{00}$$

$$N_{11} = 2 = N_{1\cdot}N_{\cdot 1}/N_{\cdot\cdot} = 8 \cdot 4/16 = N_{0\cdot}N_{\cdot 1}/N_{\cdot\cdot} = 2 = N_{01}$$

$$N_{12} = 2 = N_{1\cdot}N_{\cdot 2}/N_{\cdot\cdot} = 8 \cdot 4/16 = N_{0\cdot}N_{\cdot 2}/N_{\cdot\cdot} = 2 = N_{02}$$

$$N_{13} = 2 = N_{1\cdot}N_{\cdot 3}/N_{\cdot\cdot} = 8 \cdot 4/16 = N_{0\cdot}N_{\cdot 3}/N_{\cdot\cdot} = 2 = N_{03}$$

The final step in constructing a 3 x 2^3 design in 16 runs is to collapse the four-level variable S to the three-level matrix variable M. Collapsing is done by a many to one correspondence. The proportional frequency condition is preserved no matter what mapping scheme is used. However, the bias properties of the estimated parameters will depend on the correspondence.

Consider the two correspondence schemes shown in Table 7. These two mapping place matrices B, C, and D in the sixteen fuel rods in the manner given in Table 8. There are other possible mappings but they all give matrix patterns which are equivalent to correspondence I or to correspondence II.

TABLE 7. Two Correspondences for Collapsing a Four-Level Variable to a Three-Level Variable

I		II	
Four-Level	Three-Level	Four-Level	Three-Level
0	D = 2	0	B = 0
1	B = 0	1	C = 1
2	C = 1	2	D = 2
3	D = 2	3	D = 2

TABLE 8. Positions in the Experiment of the Low P-C Ratios Matrices for Correspondences I and II

		Correspondence I		Correspondence II	
Temp. °C	Particle	Cycle (I)	Cycle (II)	Cycle (I)	Cycle (II)
900	high	B	C	C	D
900	high	D	D	B	D
900	low	B	C	C	D
900	low	D	D	B	D
1250	high	C	B	D	C
1250	high	D	D	D	B
1250	low	C	B	D	C
1250	low	D	D	D	B

Correspondence I was chosen as the mapping scheme to complete the design after examining how the main effects were biased by the two-factor interactions for each correspondence. Each correspondence has the temperature variable and one of the matrix contrasts confounded by two-factor interactions. But correspondence I has the advantage that the least important contrast M3 is the contrast confounded. In addition, three of the two-factor interactions TM1, TM2, and TM3 are not estimable for correspondence II, while only two of the two-factor interactions TM2 and TM3 are not estimable for correspondence I. Table 9 summarizes the bias properties of correspondence I and II.

TABLE 9. Bias Properties of Correspondence I and II

Correspondence I	Correspondence II
$E[\hat{\mu}] = \mu$	$E[\hat{\mu}] = \mu$
$E[\hat{C}] = C$	$E[\hat{C}] = C$
$E[\hat{T}] = 3/2\ T + 2\ CM3 - 1/2TM1$	$E[\hat{T}] = 2\ T + 2\ CM2$
$E[\hat{P}] = P$	$E[\hat{P}] = P$
$E[\hat{M1}] = M1$	$E[\hat{M1}] = M1$
$E[\hat{M2}] = M2$	$E[\hat{M2}] = 2\ M2 + CT$
$E[\hat{M3}] = 4/3\ M3 + 1/3\ CT$	$E[\hat{M3}] = M3$
TM2 not estimable	TM1 not estimable
TM3 not estimable	TM2 not estimable
	TM3 not estimable

FLUX VARIABLE

The flux variable is a continuous variable which varies parabolically over the length of a hole in the reactor. This variable is a function of the reactor and is independent of the cycle, temperature, particle coating and matrix variables. A measurement of the flux can be made at the center of each fuel rod and varies from a low value of 6×10^{14} MeV to a high value of 12×10^{14} MeV.

To statistically control the flux influence, the range of the flux values could be stratified and used as a blocking variable. This procedure would add another variable to the factorial portion of the design and would further complicate the construction of the experimental design. I think a better solution would be to use the flux variable as a covariate with the value of the flux at the center of each fuel rod as the realization of the covariate. In the absence of additional information, I would assume that the response is linearly related to the covariate. This linear relationship can be tested because the flux varies symmetrically about the midplane of the reactor giving two measurements for each flux value. The residual error can then be partitioned into the error between duplicates and the error due to lack of fit of the model. By comparing the two error terms with an F-test, the adequacy of a linear model may be tested.

The model used to describe the vector of responses $\underline{Y}$ can be partitioned into the factorial variables and the covariate variable. Let the mean, main effect and two-factor interactions of the factorial variables be represented by the vector $\underline{\alpha}$ with the corresponding design matrix $\underline{X}$, and let $\underline{\beta}$ represent the regression coefficient for the flux values $\underline{F}$. With this notation, the model is written as

$$\underline{Y} = \underline{X}\,\underline{\alpha} + \underline{F}\,\underline{\beta} + \underline{\varepsilon}$$

where the error vector, $\underline{\varepsilon}$, is assumed to be normally distributed with $E[\varepsilon] = 0$ and $\text{var}(\underline{\varepsilon}) = \sigma^2 I$ in the customary manner. The analysis of this covariate model can be found in Linear Models by S. R. Searle (1971).

RESPONSE VARIABLE

The observed response is the proportion of particle coatings which have broken. This type of response is usually transformed by arcsin $\sqrt{Y}$ to stabilize the variance for the analysis of variance.

A potential problem with this type of response is that the number of particles in each fuel rod may be inadequate to detect any significant results among the treatments. The expected failure rates for the matrix and particle coating variables are shown in Table 10.

TABLE 10. Expected Failure Rate

	Particle Coating	
Matrix	Low Strength	High Strength
A	>1%	∿1%
B	∿1%	<1%
C	∿1%	<1%
D	∿1%	<1%

Because of the high flux in the HFIR, only 100 to 400 fuel particles can be loaded into a fuel rod. The expected breakage then is only one to four broken particles per fuel rod which would not be adequate to detect any significant differences among the treatments. The experiment is still useful, however, as a screening experiment for detecting unexpected large proportions of particle breakage.

REFERENCES

Addelman, S., "Symmetrical and Asymmetrical Fractional Factorial Plans," Technometrics, 4 (1962), 47-58.

Plackett, R. L., "Some Generalizations in the Multifactorial Design," Biometrika, 33 (1946), 328-332.

Searle, S. R., Linear Models, John Wiley and Sons, Inc., New York, 1971.

DISCUSSION: DESIGN PROBLEMS FOR THE HTGR

Tietjen: My experience has been that one nearly always gets wiped out in something that is as hot as one of these high temperature gas reactors. The elements stick together. They get broken up in analysis later. All sorts of things happen that one doesn't anticipate. My approach would be to be much less ambitious trying nothing more complicated than a 2^2 factorial. If a fourth of your results get wiped out you can salvage something from your experiment. Do replication so that you have some kind of error against which to judge the results. If you don't, your whole experiment gets destroyed. There is nothing left!

Bayne: To do replications, we will have to convince the experimenter to not be as ambitious. I am not sure how to replicate and, at the same time, answer all the questions the experimenter wanted answered. Have you been successful in convincing the experimenter not to be so ambitious?

Tietjen: I haven't been successful. The other complication is that you are treating these fuel elements with what you suppose is a constant temperature, flux, etc., but by putting in different treatments you change the treatment to the fuel elements themselves.

Gluckman: If the engineer really wants that many variables with that fine of testing, tell him to give you more rods for your test. If he is that ambitious, you can point out the trade-off of giving you more rods with some sort of assurance that he will get his test run, rather than just the two he offered.

Bayne: It still goes back to the problem of trying to convince the experimenter that he can't get something for nothing. This complicated experiment was accepted because it was able to give the information he wanted just using the two holes.

Miller: I don't think that the experiment was all that overly ambitious. I have designed and have had fuel type experiments run in reactors that were highly fractionated and were successfully implemented. It may depend upon the man you are working for. It is obvious that this man did not anticipate the problems that were inherent in running in a high flux environment.

Prairie: I am concerned about the response variables. We are thinking about the 0, 1, 2, 3, and 4 powers. Based on this you really can't learn too much about the effects of these factors, but you are hoping to do some screening. My question is, if you find the combination that gives you a particularly high proportion of fractures, then what would you do? What is the next step? If one of the combinations gave two or three times as high a number of fractures as this one percent range you are talking about, would you do some additional experimentation to verify?

Bayne: Actually, once the screening experiment was done, the engineers were planning to eliminate certain factors and add other selected variables

to complete a larger experiment. Higher density fuel particles would preclude the use of the high temperature reactor for this experiment.

Hooper: If the full experiment had worked, how long would it have taken? Have you considered using sequential procedures in the design? In an experiment that runs through very long exposures, you don't have that sort of option, but in shorter experiments you do. In a high flux environment sequential design procedures will often save you a lot of agony.

Beggs: I wonder if temperature and flux aren't highly correlated. If so, you probably ought to use both temperatures and flux as covariates.

Bayne: Yes, they are highly correlated. Actually in order to produce the temperature, each fuel rod loading was calculated based on a flux profile. In other words, the higher the flux the fewer number of fuel particles per rod. This was done to make temperature constant, i.e., like a step function. There were no thermocouples in the reactors to measure actual temperature, so it is hard to say what the actual distribution was. The temperature was controlled by the number of fuel particles. It would be difficult to use temperature as a covariate because there were no readings of actual temperatures during the experiment.

Griffing: I was going to bring up a similar question about temperature. You said that the temperature and the flux varied in a similar manner. In a particular class of 900°C, there is as much variation in temperature as there is between the 900° and the 1250° class. That is, you get as much temperature variation within the 900° class as you would get with some members within the 900° and the 1250° class. So treating them as discrete variables would not necessarily be the most valid way to go. I don't know what alternative you would have, being as limited as you are in the experiment.

Bloomfield: There is nothing that is going to protect you against total losses of your experimental data. But the general idea of setting up designs in such a way that they are robust against small amounts of missing data is starting to get some attention.

Bayne: Peter John has done some work on missing data in fractional factorial designs and his results indicate that if you do have missing data in fractional factorial designs, then your estimates are highly confounded with other parameters in the experiment.

Tukey: There is a general comment that needs to be made and that is we now have computers. Orthogonality is not the wonderful thing that it used to be. It was never that wonderful anyway. Bi-orthogonality is important. Orthogonality is a decorative way to get there. The distinction is that the way you calculate the answers from the data is bi-orthogonal to the things you don't want to find if when they happen it doesn't affect this particular combination of experimental results. And once you face up to resistant/robust techniques, nothing is ever going to be orthogonal again. As soon as you have one soft value that you treat differently from the

others, any perfect orthogonality is gone. You only need approximate orthogonality. The result quoted from Peter John has to do with using the estimate that is as close to what we would have done if we had had all the data. This is probably the wrong thing to do. Certainly, if you have a fractional factorial with missing data and you analyzed it as a regression problem, you will have no more confounding than you will have in any other regression problem.

Nicholson: There are no thermocouples in the individual fuel elements?

Bayne: No, I don't believe the experimenter placed thermocouples in the reactor. There are flux monitors.

Nicholson: So the situation then is that we have three ANOVA variables that we can set supposedly at two nominal levels each and they will stay there. We have two temperature levels that we can set based on heat transfer calculations and a theoretical flux distribution. If these are not correct, the temperatures will be at undetermined levels which hopefully fluctuate around the two nominal levels. And then we have estimated flux levels that go along with each element.

Bayne: Yes, except for the fact that they can take a flux profile before the experiment is even run and get flux values for the reactor. I think the problem is that we are talking rods or of fuel elements. There is a profile for each hole of the reactor, and we can measure these at the center of each of the two-inch fuel rods.

Tukey: You mean this is a profile before the rod is even inserted?

Bayne: Yes.

Tukey: And there is no flux measurement in the rod itself so whatever change in flux that the rod caused is unmeasured?

Bayne: True. I still would like any comments from any one who has experience with reactor designs on how they have handled the flux variable, whether they have treated it as a covariate or used some kind of blocking scheme or whatever.

Tukey: Why not use the general maxim - when in doubt, do both?

Gardiner: I handled the flux variable in the same way you intended to handle it, Chuck, in an experiment in the Oak Ridge research reactor. This was long before HIFR was built. Unfortunately, I didn't have that nice symmetric profile that you had. The calculations were a little messier, but that's the way I handled it.

Bennett: I have done several reactor experiments of this kind. In every instance, in one way or another - and not always quite as directly as you imply - flux was used as a correcting variable. Usually you can get a

computation of flux that is at least as good as the measurement you are going to get out the other end. The thing that bothers me most about the whole experiment is the fact that you have a fairly ambitious and saturated experiment which is not quite consistent with the ability to measure the 1% failure rate on a relatively small number of elements. In other words, I don't see why you expect to get very much distinction between the design factors and the measurable results.

Bayne: I think that is correct. The experimenters feared that there would not be any results coming out. They wanted to use it as a screening experiment in case the predicted failure rates were grossly in error.

Bennett: One of the essentials of the highly complicated, or highly saturated, designed experiment is an error estimate sufficiently small and sufficiently well determined to measure significance and/or establish the existence of true effects.

Bayne: The cost for these experiments is very high; it might run as high as $100,000 or more. When you are working with this kind of price tag, the experimenter feels he should get the most for his money. That is why I think he is hesitant to not be ambitious in these types of experiments.

Nicholson: Is there enough symmetry around the center of the reactor as far as the flux profile is concerned to pair things and get some sort of error measurement by putting in duplicate situations.

Bayne: Yes, there is. At least the profile that the experimenter showed me had duplicate measurements for each of the positions in the fuel rod in the reactor near the midpoint.

Prairie: I am still worried about the response. Can you do something to make some sort of stress testing where you can beef up some factors? Flux you can't beef up. Can other factor levels be used in such a fashion as to increase the fracture proportion or would that not tell you anything? I am concerned about setting up a factorial type of experiment to look for effects on the response variable. It is going to be hard to find out anything unless something very strange is going on compared to what you think is going on.

Bayne: I don't know the answer to that question. Are you talking about using more diverse levels for the variables in order to get a higher expected failure rate?

Prairie: I am thinking of digging into the fracture distribution a little bit deeper to learn what is going on. Unless there are almost thousands of replications when you talk about a 1% fracture proportion, it's going to be hard to detect any differences at all. In an experiment like this, I don't even know what it means to estimate experimental error.

Bayne: I agree. But the levels were set by the experimenter and I really don't know how they affect the particles.

Bloomfield: I gather that the expected number of breakages is really quite small. What is the response - a mass of uranium or a mass of plutonium released by the fracture? Would you leave it as a continuous response or categorize it?

Bayne: They dissolved the fuel rod and used a chlorine leach method to calculate the amount of uranium leached from the broken particles. In that sense they could calculate the weight relative to the original weight.

Bloomfield: I am wondering whether, if the mass of the pellets is given, you can convert back to an integer number of fractures. In that case presumably you would be close to a Poisson distribution and could get some ideas of error magnitudes.

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

SAMPLING THE ENVIRONS FOR CONTAMINATION

L. L. Eberhardt
R. O. Gilbert
Battelle Pacific Northwest Laboratories
Richland, Washington

INTRODUCTION

In the last decade or so, the general public has become very conscious of environmental affairs, and use of the word "ecology" no longer evokes puzzled expressions from laymen or students. The broad responsibilities of the Energy Research and Development Administration include studies of a wide range of contaminants in environments ranging from the sub-tropics to the Arctic zone. Under Atomic Energy Commission sponsorship, a wealth of information accumulated on the behavior of a variety of substances in both biological and physical systems, including much attention to "fallout" radionuclides, in the earlier years.

Inasmuch as many contaminants may exhibit a highly dynamic behavior in both abiotic (atmosphere, water, and soil) and biotic systems, a good deal of attention has been paid to various kinds of mathematical models for that behavior. In the substantial literature dealing with the "kinetics" of various substances, one is hard put to find any guidance for the design of a sampling program for the study of a trace substance (a "contaminant") in any system. About all we have observed is that most laboratory workers tend to take observations equally spaced on the time axis, but occasionally one sees plots that apparently result from the use of equal spacing on a logarithmic scale. Collections made out-of-doors seem mostly made at convenient "sampling stations," but sometimes a grid system is suggested, or stations may be concentrically patterned on a "point source." For an impression of the current state of analysis of pollution data, the reader is referred to two volumes presumably directed towards statistical aspects of such studies (LeCam, et al, 1972; Pratt, 1974).

Since there is a rather substantial body of survey sampling theory, and because contaminants are seldom found to be uniformly spread over the

landscape, it has seemed to us that more attention to sampling design might yield worthwhile benefits. Two kinds of practical problems come up immediately. One is that the process of actually locating and sampling items of the biota is rather more time consuming that most of us realize. In short, no one has yet found a way to take a random sample of the fish in a river.

The second problem is that most sample survey methodology seems to center on estimation of a mean or a total. In many cases, such an estimate seems to be useful and worth having. However, we have the uncomfortable feeling that it is really the spatial pattern of distribution of some contaminant that is of most interest. In a general way, then, this is our problem: What are the elements of a sampling theory concerned with ascertaining spatial pattern, as contrasted with present emphasis on estimating totals or means? We are particularly concerned here with the allocation of sampling effort for the purpose of determining pattern.

It can immediately be objected that we ought also to be concerned with the temporal pattern. Unquestionably this is so. We will, however, avoid that issue here for several reasons:

- The complications in dealing with space and time together are obviously a bit too much in the beginning.
- Many of the more obvious practical problems can be dealt with in a static frame of reference.
- There don't seem to be many generally accepted models for spatial pattern of contaminants, whereas there are a host of kinetic models for variations over time. Furthermore, we believe that there is a body of theory directly applicable to "sampling in time," given a particular model. Much of the recent work has been done in industrial experimentation and stems from a paper by Box and Lucas (1959). For a recent review, see Cochran (1973).

A FRAME OF REFERENCE

In order to stay roughly within a static situation, we here mostly assume that some contaminant (heavy metal, pesticide, radionuclide, petroleum, etc.) is released to the environment and builds up there. By

far the bulk of most such substances accumulates in surface soils and sediments and is redistributed by physical forces (erosion of one sort or another). In terms of total quantities, plants and animals generally will contain a negligibly small fraction, just because the biota constitute a very small mass relative to the physical environment. We can thus suppose that a really thorough sampling survey will initially include fairly extensive sampling of substrate (surface soil or sediment), followed by a study of vegetation, and finally a survey of animals feeding on the vegetation, predators on those herbivores, and so on. The fact that sampling is sometimes confined to the higher tropic levels (e.g., predators such as trout or pike) simply reflects an understandable preoccupation with hazards to people, which are often best evaluated by sampling as close to the dinner table as is feasible. Such studies serve to evaluate immediate hazards, but cannot provide the understanding required for full analysis.

Our problem can best be stated in terms of sampling soil. Conceptually the simplest scheme is deposition of some substance from a single "point source," from which emissions either stopped some time in the past, or for which the current rate of release is small relative to the quantity accumulated over time. However, one might suppose the situation not to be too different for, say, the accumulated record for air sampling devices, if the problem is posed as one of the "best" pattern for arraying such instruments.

The data can be regarded as values of some positive, continuous random variable, expressed as quantities per unit area of surface. For simplicity we assume soil cores to penetrate deeply enough so that subsampling by depth is not needed. We also sweep under the rug such problems as "instrument" (counting) errors, "laboratory errors," pooling, aliquoting, compositing and the like. In practice one must, of course, spend a lot of time trying to be sure variance components due to these sources are small relative to between-sample variances, or are accurately measured. Normally, the samples will be very small fractions of the area sampled, and thus presumably can be regarded as coming from an indefinitely large population (for variance calculations and the like). Strictly speaking, the population is finite, is defined by the coordinate system used to locate samples, and the sampling is usually without replacement, inasmuch as samples are removed from the area.

Some kinds of vegetation (mainly densely growing crops) might be regarded in the same frame of reference as soils. Most noncrop vegetation is irregularly distributed so that a unit area sample may frequently yield no individuals of a given species or group of species. Enough is known about differences in interception of particles by various kinds of plants, and about plant physiology, to dictate that simply sampling whatever vegetation is present is not suitable for any but the crudest of "surveys."

For estimating totals or means of contaminants on vegetation, one can use a two-stage sampling scheme, where the first stage is a sizable plot (quadrat) within which individual plants (elements) of the selected species are selected at random. If one is instead mainly concerned with the distributional pattern of a contaminant, it will evidently be necessary to also consider the spatial pattern of the plants. Possibly two goals might be distinguished, one being that of ascertaining pattern of the burdens of contaminant on individual plants (as in a "food-chain" study) and the other being concerned with deposition or redistribution over space (so that the plants serve mainly as collectors).

Our most recent and most extensive experience in sampling soils and vegetation for a contaminant has been in connection with a study of plutonium at the Nevada Test Site conducted by the Nevada Applied Ecology Group over the last several years. We are thus inclined to view the problem in the framework of that study, but it is, of course, of much wider scope. Presumably it extends far beyond contamination, in the sense that rational evaluation of the present "energy crisis" requires knowledge of both the total quantity and the location of our oil and uranium reserves.

It is, in fact, in the area of "geostatistics" where we have found most of the references that seem most relevant to our problem. Since we have really only just started to look at some of these sources, we can only suggest some impressions here. The main tool in common use for dealing with distributional pattern is the contour map, which shows isopleths or lines of constant value for the variable of interest. In petroleum exploration, the contours of interest may be mainly elevations of some subsurface rock stratum (since oil is found in "folds" or "anticlines" in such formations), while mining geologists will generally be interested in mapping the extent of "grades" of ore (i.e., concentrations of certain minerals).

In both fields, the observations usually are made in an irregular spatial pattern. The computerized contouring programs with which we are familiar use these irregularly spaced observations to estimate concentrations at the grid nodes of a uniform grid pattern. The grid-node values are estimated using polynomial surface fitting, moving average, or both methods in sequence. (Davis, 1973, gives an elementary discussion of the techniques in use; also see Agterberg, 1974, the colloquium volume edited by Merriam, 1970, and the recent exchange between Olea, 1974, 1975, and Akima, 1975.)

In designing a survey for contaminants in soil, it would seem entirely feasible to take the samples directly on a grid pattern, but some kind of smoothing will nonetheless ordinarily be required before contouring is possible (contours are obtained by interpolation between the estimated grid-node values). Two statistical problems then come into view. One is the well-known issue of systematic vs. random sampling (see, e.g. Cochran, 1946, Quenouille, 1949, Matérn, 1960). The second question is whether a simple systematic sample will be efficient in terms of our problem (estimating distributional pattern). The two issues are evidently inter-related, but need to be discussed in terms of "variance-laws" (see below) to approach the problem of allocation of sampling effort.

In many of the cases where vegetation is to be sampled, the irregular distributional pattern of individual plants prevents the use of systematic sampling, so that we are then immediately in a similar frame of reference to that described above for petroleum and mining exploration.

VARIANCE LAWS

Nearly all of the data that we have thus far located on contaminants in environmental samples show a strongly right-skewed distribution. Much of the data that we have studied suggests that the sample coefficient of variation ($s/\bar{x}$) is approximately constant over a wide range of concentrations. Coefficients of variation for many of the "fallout" radionuclides fall in the range of 0.1 to 0.4 (Eberhardt, 1964; Remmenga and Whicker, 1967). In reviewing data on DDT concentrations we found half of 74 sets of data (mostly quite small samples) exhibited coefficients of variation between 0.3 to 0.5. Gilbert, et al, (1975) tabulated coefficients of variation for 80 sets of data (sample sizes 5 to 50) obtained in stratified random sampling for

plutonium 239-240 in soil and vegetation at the Nevada Test Site. The modal value is 1.0 for both soil and vegetation, while the median for soil is somewhat higher (about 1.2) but lower (about 0.9) for vegetation. Some limited data for petroleum in the marine environment suggest values of the coefficient of variation to be roughly unity for that substance (National Bureau of Standards, 1974).

Frequency distributions commonly used as models for this kind of data include the gamma, lognormal and Weibull distributions. For plutonium, at least, the distribution of particle sizes no doubt plays an important role in the extreme variability observed in both soil and vegetation samples. With the exception of the plutonium data, most of the studies summarized above were based on rather uncertain sampling schemes. For this reason, and in consequence of the strong skewness and small sample sizes available, we do not suppose there is much point in trying to decide on a specific frequency distribution model. The important points seem to us to be the substantial skewness of the data and approximate constancy of the coefficient of variation over a wide range of concentrations. From this we suppose it to be desirable to consider a "variance-law" of $\sigma^2 = \gamma^2\mu^2$ where γ is the coefficient of variation.

It should be noted that the plutonium data exhibit a very marked spatial trend, crudely concentric about the "ground zero" of the "safety-shots" giving rise to much of the data (cf. Gilbert, et al., ibid.). We ordinarily used four to six strata to reduce the effect of the sharp gradient on between-sample variability, but obviously this "trend-effect" nonetheless contributes to the within-stratum variances from which the coefficients of variation were computed. As yet we have not had time to try to sort out a trend component, but there seems to be ample evidence that a substantial part of the variation observed within strata is <u>not</u> due to trend.

SOME ALLOCATION SCHEMES

When some prior or auxiliary information is available for establishment of strata, three schemes are commonly considered for allocating a fixed total sample (n) among strata. These may be listed as equal, proportional, and optimal allocation. Equal allocation might be considered

for "analytical" comparisons (i.e., testing for significant differences between stratum means), if one assumes the underlying frequency distribution to be lognormal and transforms accordingly, and if no particular set of contrasts is deemed more important a priori. Cochran (1963: 145-146) deals with allocation for comparisons between strata (or "domains of study") when within-stratum variances are not equal. Quite reasonable arguments might be raised for dealing with the problem proposed here in terms of comparisons between subareas, but we have viewed it entirely as a matter of location--"where is the contaminant?"

If A_h denotes the area of the h^{th} stratum, and $\sum A_h = A$, then in proportional allocation we have:

$$n_h = \frac{nA_h}{A} .$$

In optimum allocation (assuming the finite population correction negligible, and assuming cost per unit is the same in all strata) we have:

$$n_h = n \frac{A_h S_h}{\sum A_h S_h} = \frac{A_h \mu_h}{\sum A_h \mu_h} ,$$

where the right-hand quantity results from assuming the variance law above, i.e., $S_h = \gamma\mu_h$. It is interesting that optimum allocation thus amounts to allocation in proportion to the fraction of the total quantity of the substance falling in the stratum.

Another candidate for an allocation scheme may be postulated by supposing each stratum to be comprised of a number of subareas or cells, all of the same area. If it is required that the quantity within <u>each such cell</u> be estimated within a common, constant standard error, δ, then:

$$\delta = \frac{S_h}{\sqrt{n_{jh}}} = \frac{\gamma\mu_h}{\sqrt{n_{jh}}} \quad \text{and thus} \quad n_{jh} = \left(\frac{\gamma}{\delta}\right)^2 \mu_h^2,$$

where n_{jh} is the sample size required within each cell of stratum h, and thus n_h is proportional to $A_h\mu_h^2$. Consequently the several allocations proceed as proportional to A_h, $A_h\mu_h$, and $A_h\mu_h^2$. The last scheme might also be regarded as a restricted randomization, or as two-stage sampling.

One immediately evident difficulty with systematic sampling under the variance law mentioned above is that systematic sampling presumably corresponds roughly to proportional allocation, and thus may not offer sufficient precision in high concentration areas while "oversampling" the low concentration regions. We are thus tempted to propose systematic (grid) sampling within strata, but using a variable grid mesh, so that the strata having high concentrations have a fine mesh, while those with the lowest concentrations have a much coarser mesh. However, it is not clear that such a scheme will be compatible with the available contouring algorithms (we expect trouble at stratum boundaries).

While the contouring schemes offer a neat graphical representation of concentration data, it is by no means certain that contouring is the best way to deal with the problem posed here. Two-stage sampling with stratification of the first stage units may well be a desirable alternative, and perhaps our "problem" only expresses a need for better definition of objectives. Certainly this has been a major problem for us in trying to help investigators and administrators decide on survey designs.

CONCLUSION

Our purpose here has been to express a concern about surveys for contaminants that focus on estimation of means or totals, and to suggest a need for some theoretical and practical emphasis on sampling for distributional pattern. We have postulated a static system to avoid the ramifications introduced by kinetic models and to direct attention to spatial patterns. In effect, we are assuming a system that changes rather slowly over time, and we presume that a theory suitable for our "problem" will ultimately need also to be examined for its relevance to patterns of change over time.

It should be emphasized that our main present concern is with sample allocation, a priori. In the situation where one is given a set of observations and asked to produce a map, then it may be mostly a matter of judgment as to which of the several available schemes for data smoothing are used.

REFERENCES

Agterberg, F. P., Geomathematics, Mathematical Background and Geo-Science Applications, Elservier Scientific Publishing Company, New York, N.Y., 1974.

Akima, H., Comments on "Optimal Contour Mapping Using Universal Kriging" by Ricardo A. Olea, J. Geophys. Res, 80, No. 5 (1975), 832-834.

Box, G. E. P., and Lucas, H. L., "Design of Experiments in Non-linear Situations," Biometrika, 46 (1959), 77-90.

Cochran, W. G., "Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations," Annals of Math. Stat., 17 (1946), 164-177.

Cochran, W. G., Sampling Techniques, J. Wiley and Sons, New York, (second edition), 1963.

Cochran, W. G., "Experiments for Nonlinear Functions," J. Amer. Stat. Assoc., 68, No. 344 (1973), 771-781.

Davis, J. C., Statistics and Data Analysis in Geology, J. Wiley and Sons, New York, 1973.

Eberhardt, L. L., "Variability of the Strontium-90 and Caesium-137 Burden of Native Plants and Animals," Nature, 204 (1964), 238-240.

Gilbert, R. O., Eberhardt, L. L., et al., "Statistical Analysis of $^{239-240}Pu$ and ^{241}Am Contamination of Soil and Vegetation on NAEG Study Sites," Radioecology of Plutonium and Other Transuranics in Desert Environments, Nevada Applied Ecology Group Progress Report as of January 1975 (White, M. G. and Dunaway, P. J., Editors), USERDA NVO-153, 339-448.

Lecam, L. M., Neyman, J, and Scott, E. L. (editors), Effects of Pollution on Health, Vol. VI of Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Univ. of California Press, Berkeley, 1972.

Matérn, B., "Spatial Variation," Medd. fr. Statens Skogsforsknings Institut, 49, No. 5 (1960), 1-144.

Merriam, D. F. (editor), Geostatistics, Plenum Press, NY, 1970.

National Bureau of Standards, Marine Pollution Monitoring (Petroleum) Symposium and Workshop, National Bureau of Standards Special Publication 409. U.S. Government Printing Office, Washington, DC, 1974.

Olea, R. A., "Optimal Contour Mapping Using Universal Kriging," J. Geophys. Res., 79, No. 5 (1974), 695-702.

Olea, R. A., "Reply [to commentary by Akima]," J. Geophys. Res., 80, No. 5 (1975), 835-836.

Pratt, J. W. (editor), Statistical and Mathematical Aspects of Pollution Problems, Marcel Dekker, Inc., New York, 1974.

Quenouille, M. H., "Problems in Plane Sampling," Annals of Math. Stat., 20 (1949), 355-375.

Remmenga, E. E. and Whicker, F. W., "Sampling Variability in Radionuclide Concentrations in Plants Native to the Colorado Front Range," Health Physics, 13 (1967), 977-983.

DISCUSSION: SAMPLING THE ENVIRONS FOR CONTAMINATION

Easterling: Was the laying of the grid done after the contaminant was released? How about experiments where you don't mark your contaminant deposition as it happens? Have you looked at that or are you going to provide us with references? I am thinking in terms of a nuclear power plant. If we want to monitor a possible effect on fish life, how do we place stations in the river or the lake? I was wondering if there was any related work that you looked at that might be of help.

Eberhardt: We worried about that (power plants), too. I think it would be a whole lot harder in the sense that the kinetics of what happens over time has to be considered. I did mention in the paper some schemes that have been developed, in industrial experimentation, to guide sampling in time or sampling of some continuous variable that influences the outcome of the response. Those things undoubtedly need to be looked into for just that kind of problem.

Here, we would like some discussion relative to a theory of sampling for distributional pattern, given nonhomogeneity of variance and some prior information. We would like to contrast it with the usual designs which seek to optimize estimates of totals or means. In other words, "What are the distinctions between sampling for pattern and sampling to estimate how much there is?" Such sampling is not so difficult if the variance is homogeneous over the field. Then, I would settle for systematic grid patterns. If the variances aren't too bad, one can go to the contouring routines and come up with a nice map. I suppose it should be said, too, that we don't know that the contour scheme is "the" answer. It is just that it is a neat description of the pattern that we infer from the data we collect. There may be some other entirely different way of doing it.

Ferris: Rocky Flats has been doing experimentation and data analysis, even some experimental design, in soil sampling for quite some time. Our situation is somewhat different. We do have a point source. We pretty much know how it was dispersed - mostly through the wind. We have covered all fronts in that we use both a systematic grid as you showed and, in case that doesn't work, we also have a polar design. Don Michaels from Aerojet has been doing a lot of work on grid sampling of the soil. In fact, he designed our standard sampler that we use at Rocky Flats. As I understand it, it is patterned after the one that the Colorado State Department of Health uses. We have many people monitoring as well as ourselves - the State Department of Health, the universities, and interested ecology groups. Relative to movement of plutonium in the soil, John Minor at Rocky Flats is doing a lot of work and is publishing on how plutonium gets from here to there in the soil. Does it go radiologically, does it dissolve, or does it get carried along with another particle? I am not talking about linearly; I am talking about going down into the earth. You might find his work interesting also. Another thing I have noticed is that, in all soil work, the log normal distribution is used. I don't know

why. It seems to be convenient. I hear the story that it is used because it is there and that is what has always been used. I wonder if anybody has done any work to see if there is another distribution that is better. Is this something that is empirically derived or do we just know that plutonium in soil is distributed log normally?

Crain: We at Lovelace are interested in distributions of plutonium and other radionuclides, but at a different level. We are interested in the distribution of spatial patterns of inhaled particles. When the investigators came to discuss this problem with me in connection with alpha emitting particles last spring, I looked at what had been done by the ecologists. I rummaged through their literature and found quite a bit of material. Foresters are interested in problems of spatial pattern and also problems of density of estimation, which seems similar to the kind of problem you are getting into of not only total inventory but the spatial pattern. The reason we are interested in it is that it has to do with microdosimetry problems and the hot particle problem. But ultimately we would be interested in the question of time because we feel that there is a possibility that either inhaled particles are annihilated in a specific way or are moved around by some sort of clearance process. To return to your problem, it would be interesting to look at the size distributions of these particles. You can probably get some of that from your autoradiograph studies. I think it would be interesting to relate those to some kind of theory about how they might be blown around. Perhaps the heavier particles are not distributed the same as the lighter ones.

Eberhardt: The problem in forestry or plant ecology generally has to do with discrete individuals that are readily identified, readily seen. Here we must take a sample and send it to the laboratory for analysis, and even then we have "counting" errors and various biases to contend with. But I agree there are very similar sorts of angles that are indeed interesting. We have a related situation on the Test Site because we are very much interested in the relationship between the concentration in soil at a spot and the concentration in vegetation near that spot. Those two factors certainly are important in human exposure problems.

As far as the application of the log normal distribution is concerned, we worried about that several years ago. We contrasted the gamma and the log normal as possible models. We did not look at the Weibull, which seems to be another possibility. We found that it would take tremendous samples to distinguish the two distributions. The big problem of course is in the tails of the distributions. We simply don't have samples adequate to distinguish between such similar distributions. We tried various tests for kurtosis, skewness, and so on. We then went to simulation. If we considered the gamma and log normal as alternatives and decided to use one and picked the wrong one, it didn't make much difference in the outcome. Log transforms of the gamma, interestingly enough, behave pretty well. That's mostly approximation by simulation, but partly built on some other early work with tests for homogeneity of variance by Barlett and others.

Bloomfield: The gamma distribution with coefficient of variation one is the exponential distribution, and the logarithm of an exponential distribution is pretty heavily skewed. A gamma with larger shape parameter necessitates a smaller coefficient of variation.

It is unfortunate that our maps use a rectangular grid system. Barocentric coordinates would be better since a hexagonal grid gives more uniform coverage than a rectangular grid. But I guess it is hard to get people to go out to barocentric coordinates and take samples.

You mentioned that Kriging is different from moving averaging. It is different in that the weights in the averages are determined from the semi-variogram, rather than being determined on some ad hoc basis. But it still shares the moving averages property of being a linear function of the data. Therefore it is sensitive to bad values. The technology that John Tukey mentioned earlier of adaptively weighting things based on the size of residuals can be applied to Kriging to get resistance to the occasional bad value, or generally long-tailed errors. You hear it said of Kriging that it allows for an infinite or nonconstant variance. That is absolutely true. But it is still true that it is a technique that is designed for use with Gaussian distribution. The infinite variance part should not lead one to beleive that it is good with stably distributed Cauchy errors or such. It is still fundamentally a procedure designed for Gaussian data.

When you applied one of the smoothing procedures to the raw data before taking the logarithm, the gross disparity in the sizes gave you a rather uncomfortable amount of smearing out of the peak. That gave a strongly sloped residual versus smooth plot. If you fit a straight line to this plot, you get a function that says residuals tend to look like a linear function of the smooth values. If you then subtract that from the residuals, which is equivalent to adding it to the smooth values, you get a new smooth value which necessarily gives you a zero slope in that particular plot. It might remove all the smoothing. I don't know. But it does get rid of the slope.

Eberhardt: We haven't tried it (smoothing) and it is worth looking into. I do think we should say again that if we transform to logarithms then things smooth out and look pretty nice. It's just that we will then have to talk to people in terms of logarithms which we are not always successful in doing.

Steck: It occurs to me that if you are satisfied with the contours then you might design your experiment just to find particular contours. If something like stochastic approximations exist in two dimensions, then you could zigzag around on a contour and define it and then go on to a different contour.

Eberhardt: That's an interesting idea. One of the things that we need is a situation where the variance is not so unreasonable, where the sampling is not so expensive, so we can get more experience with larger samples.

Tietjen: Perhaps the ultimate objective in looking at this is to determine where it is so we can decontaminate the area where it is located. With that in mind, I would think that in the peripheral area, for instance, one would say there is an upper threshold number of counts in that first contour. Consequently, suppose we do not want to investigate that area at all, but we decide that the close-in area, inside a specific contour, will have to be cleaned up. If we think that there is more plutonium in the middle area than in the peripheral area, we perhaps should just move toward the source until we get to something that we decide is intolerable. We won't investigate the extremes at all but simply see where the high contamination begins and where it ends, i.e., discover what part has to be cleaned up. Is that an objective of the ultimate experiment?

Eberhardt: Yes. That is one of the objectives. I hesitate to take that as a primary objective and to sample accordingly. One of the troubles is that people have to decide to what degree they want to clean up an area. They would need to know the concentrations over the whole field to find out how much dirt they would have to pick up. How many tons of dirt would have to be moved to reach the 100 pci/g level, for example? You really need the whole picture before you make your decision. If you are given a guideline that says go down to a certain level and that is the sole purpose, fine. Suppose we go ahead and make a decision to remove soil out to a specific contour. We clean up the dirt out to that point and take it away. The first question then asked is, "What contaminated soil did we miss in the removal process?" We then have rather a different sampling problem, I am afraid. One of the reasons I am worried about the grid pattern is that it appears that the machine, perhaps a grader, that we use to clean up an area will probably windrow the dirt. I am concerned that, when I am now told to come back and again sample this cleaned-up area, those windrows may cause problems. You can take out dirt once and go back, do another survey and find that you didn't get it all and go back again, and so on.

Ferris: You are talking about only taking out the contaminated soil. Perhaps a better concept would be to control the contamination. If you blade it and then scoop, you hope that you will do that on a nonwindy day. Otherwise the people downwind may not be as enthusiastic about your cleaning up as you are. Let's think more in terms of locating and controlling it. If the best way to control it is to blade it, scoop it, and bury it or whatever, fine. Otherwise, an asphalt path over it would be better.

Hooper: It seems important, particularly where you want to try to estimate distribution, that in deciding on a sample plan you characterize the

source term. In a more complicated situation like Rocky Flats, substantial amounts of material that you might expect to find there are the result of a single discrete incident in time, as opposed to material that was distributed there from day to day operations. If you conclude, based upon preliminary analysis, that the bulk of material is due to an accident of some kind, then the micrometeorology at the point of that accident would be pretty compelling as far as a sample plan is concerned.

Eberhardt: We do have advance data on the site we have studied, collected with field instruments, which are insufficient to do the final work but do give us information on the pattern. We would also like to know, of course, much more about distribution, as you are suggesting.

Martz: You mentioned earlier in your presentation one of the problems with the moving average isopleths construction procedure is the inability to get a handle on the variance associated with the isopleths themselves. Is that correct? Once enough data are obtained from a certain portion of the grid to establish one of these isopleths, is it then possible to do sampling along that isopleth to verify the quality of the fit? Do you then modify the fine line that you have with some sort of a band or some sort of a fuzzy wide line to represent the error associated with these contours? Sort of sequentially sampling in the sense of establishing the accuracy of the isopleths fitting themselves?

Eberhardt: That is not something we have tried. It apparently is part of the geologists approach to get some sort of confidence measure. What they come up with is a map showing the relative variability over the field.

Griffing: I am not sure about the validity of this next idea. I am suggesting the possibility of coming up with an initial density to use in sampling. It may be based on data you already have or on some theory as to how the particles might be released. If it comes from data, it would be rather grainy like the data you presented earlier. In that case you would smooth to get a density function. Then you would do a random simulation of a grid based on the density function that you get from this initial knowledge or from an idea as to where your particles would be. In the areas where you thought you would have a larger number of particles you would be just shrinking down your grid. You wouldn't be using any kind of arbitrary grid pattern like a polar grid or rectangular grid. But the grid spacing on the average would be proportional to the expected amount of particles at any particular point.

Eberhardt: It is intriguing to think about varying grid meshes over the field. I have no idea how I can translate that into a real sampling scheme. One thing that intrigues me about simulating the situation is that it constitutes a way to get some experience with various schemes. As yet we know almost nothing about the possible auto-correlation function. There is a marked trend, which is this hill-shaped concentration pattern. The very substantial variation about this pattern possibly makes a "signal".

But as to whether there really is any relationship other than this gradual to steep change in concentration between adjacent samples I do not know. Probably not in this case.

Lohrding: I have wandered around on that test site. One thing I noticed was that in a lot of areas you have problems where there is very little vegetation. Where there is vegetation you tend to have soil building up under the vegetation. Is it possible that a lot of the heterogeneity in your data is due to the fact that you are close to sagebrush? Concentration of plutonium in pockets could introduce a lot of the variance to your data in local areas. In other words, maybe your measuring techniques are too good and yielding data which are too good. Perhaps you need to be up in the air about 100 feet and be averaging over a much broader area. Then maybe the plots will be much smoother.

Eberhardt: You are quite correct. There are little "blow sand" mounds that appear under the bushes and there are differences in concentration associated with these features. As far as we can tell, these are not major differences. That is one of the things, again, that we swept out of sight so as to get to the central issue here.

Conover: It is the central issue that I was wondering about. If your central issue is one of inventory, as you have originally stated, then it seems natural to want to sample much more heavily in areas where concentration is higher, much more so than what was indicated. Also it would seem natural to give the person collecting the sample a certain amount of leeway in the way of, say, wild card samples. As he goes from one specified spot to another, if he finds a likely site for high concentration he should have the authority to take an extra sample now and then and mark in some way where this was obtained in an effort to obtain an accurate inventory. Finally, as was pointed out, he should state the covariates that are available. This is cheap information. What is the terrain like where he obtained his sample? Are there some peculiar characteristics that would indicate that this might tend to be an area of high or low concentration? As long as this is cheap information, why not write it down?

Eberhardt: Covariates we have used are in the form of FIDLER instrument measurements. It is not quite good enough to do a double sampling kind of job, i.e., to use a cheap field instrument and expensive laboratory determinations together to get a better inventory estimate. There is an intermediate stage in which we use a laboratory crystal to make a quick determination rather than going through the chemistry. So we have looked at double-sampling, but again I put it aside for the presentation here. I would indeed like to see a good theory of adaptive stratified or other optimum sampling for inventory. I won't, from bitter experience, let the field man make decisions in quite the way you have suggested, but I think we should consider that possibility in other contexts.

Let me reinforce the inventory point. We use an objective of "inventory" because that is the one loud and clear thing we could get from the people

who were responsible for this whole effort. We used stratified random sampling with optimum allocation which does indeed put most of the effort in the high concentration areas. But, as an afterthought, we began to wonder, "Isn't there another way to look at this?" What I would like to see is a clear cut theory for sampling for pattern on one extreme, with the existing theory of survey sampling on the other side, because I know full well that in the natural field situation we are going to have a mixed bag of objectives and we will be forced to compromise. But it would be nice to look at this other business as a model on one extreme to help in judging the right level of compromise.

Zeigler: I am curious because of an incident that happened not too long ago at Los Alamos. We found a few particles of plutonium over in the townsite area that was possibly from some time long ago. As a consequence of this, we were surveyed by a low flying helicopter with on-board detection equipment. I am curious if such equipment has been used in either the Nevada test site or Rocky Flats. This is a very good smoothing type of instrumentation. When you go over these areas, it tends to smooth the data because you are seeing an area on the ground, rather than a point source. Secondly, if such surveys have been made, how do they correlate with the ground data? I would think that in the long run it would be much cheaper and much more desirable to do an airborne survey than a ground survey.

Eberhardt: Airborne surveys have been used at the Test Site and at Rocky Flats. There are all sorts of intermediate stages, starting with our FIDLER instrument, moving up to a much larger field integrating instrument, and on to the aerial array. All of these integrate and show nice contour patterns. I can't answer accurately as to the correlations between those and the ground samples where we took dirt and did a chemistry analysis. For the FIDLER, the correlations were good but not good enough to be the only thing we looked at before predicting inventory. Certainly we need to look more at the interrelationships between various scans.

Gilbert: The FIDLER integrates and is held a foot above the ground. The correlations, which Eberhardt indicated were not very good, are between the FIDLER and a 10 gram sample of soil located underneath the FIDLER's crystal. In a hot particle situation you wouldn't necessarily expect a very good correlation with that configuration. I am not quite sure what we want to look for there. Just because you don't get a good correlation doesn't mean you can't get a reasonably good estimate of the surface.

Eberhardt: Technically, the particle problem is an important facet of the particular problem we are talking about here, i.e., surveys for radiation. It should eventually be brought into the theory we are talking about here.

Zeigler: The detection equipment that I am familiar with use mostly gamma type scans. Gammas penetrate quite a distance through the atmosphere and the energy levels can cover a wide range. Some of the detection equipment that we have looked at have airborne computers with analog-to-digital

devices on the aircraft itself. You can get a complete spectrum on that. With a helicopter, you can hover if you have to.

Ellesson: It seems to me we are falling victim to the tyranny of exactness. I don't think you really need to know as much as our comments indicate. A very coarse grid may be more than you will need to develop the kind of contours that will allow the field man to make the type of decision he is going to make. You don't have to have every little wrinkle in there. He is not going to direct the bulldozer to leave that sagebrush and take this one. He is going to go out there and cut a very broad swath. Unless you are estimating the total amount, the proportional Neymann allocation of samples may not be applicable to distributional estimating since they were designed for estimating mean values. As soon as you leave estimating procedures like that, stratification comes under some question.

Eberhardt: I would like to comment on the need to be careful in our use of the word "cleanup". We were talking here as if we were actually cleaning up an area. What we are doing is taking dirt containing plutonium, hauling it someplace else, and perhaps burying it a little deeper. If we must remove plutonium from the surface soil in a particular place, I would like to do that with a very minimum of dirt that would create a problem somewhere else. There is a need for real accuracy in our work.

Martz: One thing we talked about a moment ago is alternative ways to analyze data that we might get from an experiment such as this. In regression analysis, if I had two design variables x_1 and x_2, and sample points over a grid in two dimensions on the surface, I would consider that to be a pretty hot dog experiment. Are there any advantages to using regression analysis as a tool in fitting surfaces to describe the amount of observed plutonium? One of the problems that we talked about earlier with moving average procedures is that only neighboring points give rise to the surface. In regression analysis you have all data points supporting each other in some sense plus the other advantages of variances and so forth. Has there been any experience with drawing these contours from fitted regression curves based on this much data? You would think that they would be pretty good.

Eberhardt: We should say we don't have an absolute measure of either quantity. If you think of regression in the sense of chemical determination of soil plutonium, which has error, and the radiological reading with a field instrument, then I might rephrase what you are saying in the sense of a linear combination or a weighted combination of the two sources, one much cheaper than the other. That is an interesting idea. We had used regression in the sense of double sampling for inventory.

Martz: I was just thinking of simply regressing the observed amount of plutonium versus the longitude and latitude or whatever the coordinate measures are for location of where the sample was taken.

Eberhardt: You run out on a polar ray. It's worth looking into.

Tukey: I will begin by responding to a little of this last discussion. Our lovely six-degree polynomial that went swooping up into the sky is what happens when you regress things on simple functions of x_1 and x_2. The good thing about the local methods is that they are local and you can't be ruined by irrelevant data somewhere else. And if you don't know what the shapes of things ought to be, fixing the functions in advance is likely to get you in just that trouble.

I have about three different sets of things to say. One of these involves the pattern problem. I would certainly want to work on a logarithmic basis. I don't see any excuse for not using the logarithm. It seems to me that most of the decisions you make go that way, the chemistry goes that way, the finite part of the biology goes that way, and maybe even some of the physics goes that way. Nothing goes in linear concentration that I know of except inventorying to find how much total is there. This is mostly a minor problem and certainly is not very often a high precision problem, unless you are supposed to inventory how much is there and then determine whether one percent or three percent blew away. In which case you have my sympathy.

I think we need to make a distinction between the processes of smoothing or, if you have data at irregular points and are recovering something at the good points, which is really a smoothing procedure. We're separating that from contouring. You can deal with them somewhat separately. There is a very major issue here about a repetitive approach. For anything that is as vigorous and as relatively well behaved as what you have here, it seems to me that it is very different than some of the things I have seen in geological mapping. It is wrong to expect anybody's procedure to work on a once-through basis. What you want to do is go once through, smooth that, and then come back and work on either the difference or the ratio. If you are in logarithms you would, of course, use difference. In actual units, you would almost certainly use ratios, which is another reason for taking logarithms. In other words, now you are in a place where you can do your choice, say two times.

Let's consider another major point; i.e., the FIDLER versus the analysis. You were saying that you didn't think that the FIDLER was good enough for the double sampling. Now I am not quarreling with that if you want to do inventory. But if you want to do pattern I think you missed a very large opportunity. What is the relative cost of FIDLER measurement versus an analysis? I would argue that the efficient way to do this sort of thing would be to go out and do the FIDLER measurements and smooth the FIDLER measurements in a good way. That is going to require repetition. Otherwise you pull the very skirts up as you are very well aware. Then go into the field where the purpose of the wet analysis is to get the ratio of the analysis to the smooth FIDLER measurements. Let me do a one dimensional version of this. We go out and we get a lot of FIDLER measurements and

they look as shown in Figure A. I am on a log scale so they don't look so horrible as they would otherwise and they aren't as horrible. I will smooth them with some basis or other.

I now go out and get a few analytical measurements such as the dots in Figure A. I should be looking at how these compare with the smoothing I have of the FIDLER measurements. So I smooth the analytical measurements (Figure B). Then I can put the two together. Instead of using the FIDLER as a covariate afterwards, I am essentially using it as a reference base from which to make my measurements. As long as this is a good set of samples and I have a decent process for reducing things, I can take advantage of everything the FIDLER measurement has told me about the structure. Now when I come back, I probably will have to deal with things repetitively again. If you have anything decently structured, you cannot expect anybody's analytical techniques to do a good job on one pass, with the possible exception of some Kriging techniques.

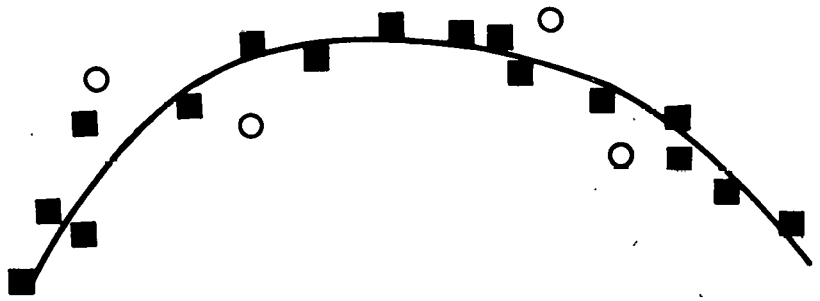


FIGURE A

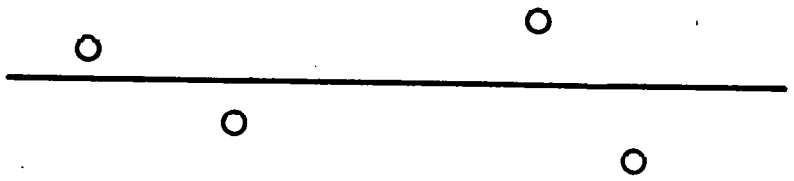


FIGURE B

On the question of grids, I don't see any reason why I shouldn't use an elliptical polar grid here. These are supposed to be concentric ellipses whatever they look like. When I want to contour, I am going to be perfectly happy to contour as if it were rectangular, subject to the fact that I have gone around the circle by going on out and repeating the measurements on each side. I am perfectly willing to use the surface program on polar

or elliptical grids, tell it is a rectangular grid and tell it to go and do the best it can. And then I can always make a change of coordinates. It seems to me that I have to expect to make a profit out of doing this. The distances that naturally come into most of these programs ought not to be ground distance. They ought to take account of what is going on. If from FIDLER measurements you think an elliptical thing is going to be reasonable, then you have to say that moving across concentric ellipses is moving a long way but moving around is not very much. Instead of drawing ellipses, what I should probably be telling the contouring device is that I have got circular patterns. This gets repeated on each side as I go round and round the circle and the distance around the circle is short, but the distance out radially is long. I want things to be analyzed in that way. If I go to random allocation within strata or within cells or something like that, I have irregular locations. I want to get that washed out. The nature of the atmosphere, for example, is very anisotropic. Things in the atmosphere change much more rapidly going up than going sideways. If you want to divide the atmosphere into cells, they have to be shaped like pancakes. I am saying that if you want to divide this area into cells they would tend to be long pieces out of elliptical annulae. That is the sensible way to divide it in terms of what you want to understand about the things from the preliminary approach. Let's cellulate it that way and take advantage of all we know. And also let's make use of a preliminary fit and work with the corrections to the preliminary fit. The preliminary fit doesn't have to be right. As long as it is better than a constant you will probably do a better job of analyzing things overall if you deal with the differences from the preliminary fit. I am sure that what the smooth FIDLER values will give you is a darn sight better than a constant in terms of predicting what the plutonium situation will be.

A word or two about Kriging. There exists two books by a Frenchman named Matheron. I can't quote the titles. But that ought to be enough to get you moving. Don't take the structure of these books too seriously. It reflects the structure of French science and not the true state of affairs. If you are to be a legitimate scientist in France, you must have lots of theory and high powered mathematics. So you will find these books talk about higher order stationary processes and so on for 12 chapters. In the last two or three chapters you get some of the meat. I think the universal Kriging more or less is following Matheron's lines, but don't feel that you have to have the high powered machinery in order for it to be useful to you. If you have irregular points, then let me require that I will not make an error if these points lie perfectly on a plane. If the data points lie perfectly on a plane, I would not like to make an error. Subject to that and subject to what the semivariogram says on how differences go with distance, then I will try to do the best I can. But saying that I require the planes to behave properly, or conceivably quadratic surfaces, takes an awful lot of points. I am now allowing for large scale trends which are not in any sense stationary and separating these slow changes that can't be stationary in many problems. They can't be stationary in yours. You've got one high hill and there aren't any others around. But

it might be that, if you take out that high hill, the other variations, particularly in the logarithm, might be reasonably stationary. The variogram plots an average of $y_i - y_j$ against the distance between these two things. It's a common thing to find situations that behave like this - where the variogram does not start from 0. This is the sort of hot particle problem. The jump at the origin is the contribution to variance that you get if you take another 10 gram sample next to this one. And that may be fairly large. It seems to me the correct picture here is the variogram picture that is associated with one of your hypotheses.

LIST OF ATTENDEES

LIST OF ATTENDEES

Janet Anderson
Argonne National Laboratory
Argonne, IL

Kenneth Apt.
Los Alamos Scientific Laboratory
Los Alamos, NM

E. E. Ard
Sandia Laboratories
Albuquerque, NM

Fred J. Balkovetz
Aeroject Nuclear Company
Idaho Falls, ID

C. K. Bayne
Union Carbide Corporation
Oak Ridge, TN

J. J. Beauchamp
Union Carbide Corporation
Oak Ridge, TN

R. J. Beckman
Los Alamos Scientific Laboratory
Los Alamos, NM

William Beggs
Bettis Atomic Power Laboratory
W. Mifflin, PA

Thomas Bement
Los Alamos Scientific Laboratory
Los Alamos, NM

Steven J. Bengston
Aeroject Nuclear Company
Idaho Falls, ID

Carl A. Bennett
Battelle
Seattle, WA

Peter Bloomfield
Princeton University
Princeton, NJ

K. O. Bowman
Union Carbide Corporation
Oak Ridge, TN

Lyle D. Broemeling
Oklahoma State University
Stillwater, OK

John Brolley
Los Alamos Scientific Laboratory
Los Alamos, NM

Larry Bruckner
Los Alamos Scientific Laboratory
Los Alamos, NM

Ray L. Buschbom
Pacific Northwest Laboratories
Richland, WA

Bill Buzbee
Los Alamos Scientific Laboratory
Los Alamos, NM

Katherine Campbell
Los Alamos Scientific Laboratory
Los Alamos, NM

Kamal C. Chanda
Texas Technological University
Lubbock, TX

Jesse Cheadle
Los Alamos Scientific Laboratory
Los Alamos, NM

C. W. Childers
Sandia Laboratories
Livermore, CA

Boyer Chu
EPRI
Palo Alto, CA

Elon Close
Lawrence Laboratory
Berkeley, CA

W. J. Conover
Texas Technological University
Lubbock, TX

Chester R. Crain
Inhalation and Toxicology Research
Albuquerque, NM

Gary L. Crellin
GE-FBRD
Sunnyvale, CA 94086

James M. Davenport
Texas Technological University
Lubbock, TX

C. J. De Carli
Sandia Laboratories
Livermore, CA

Benjamin S. Duran
Texas Technological University
Lubbock, TX

R. G. Easterling
Nuclear Regulatory Commission
Washington, DC

L. Lee Eberhardt
Pacific Northwest Laboratories
Richland, WA

J. R. Ellefson
Sandia Laboratories
Albuquerque, NM

Yvonne Ferris
Rockwell International
Golden, CO

F. Andrew Ford
Los Alamos Scientific Laboratory
Los Alamos, NM

Jerome Friedman
Stanford Linear Accelerator Center
Stanford, CA

D. A. Gardiner
Union Carbide Corporation
Oak Ridge, TN

Donald P. Gaver
Naval Postgrad School
Monterey, CA

Ethel S. Gilbert
Pacific Northwest Laboratories
Richland, WA

Richard O. Gilbert
Pacific Northwest Laboratories
Richland, WA

Leon Gilford
Energy Research and Development Adm.
Washington, DC

Perry Gluckman
Lawrence Livermore Laboratory
Livermore, CA

D. G. Gosslee
Union Carbide Corporation
Oak Ridge, TN

M. A. Griesel
Sandia Laboratories
Albuquerque, NM

Daniel E. Griffing
Atlantic Richfield Hanford Company
Richland, WA

I. J. Hall
Sandia Laboratories
Albuquerque, NM

Judy L. Harris
Pacific Northwest Laboratories
Richland, WA

Richard L. Hooper
Pacific Northwest Laboratories
Richland, WA

Marcy L. Hopper
Argonne National Laboratory
Argonne, IL

Myrle Horita
Los Alamos Scientific Laboratory
Los Alamos, NM

M. Hume
Rockwell International
Golden, CO

Ronald Iman
Sandia Laboratories
Albuquerque, NM

James W. Johnston
Pacific Northwest Laboratories
Richland, WA

David Kahaner
Los Alamos Scientific Laboratory
Los Alamos, NM

Samuel Kao
Brookhaven National Laboratory
Upton, L.I., NY

W. E. Lever
Union Carbide Corporation
Oak Ridge, TN

Ronald Lohrding
Los Alamos Scientific Laboratory
Los Alamos, NM

Victor W. Lowe
Los Alamos Scientific Laboratory
Los Alamos, NM

Warren Lyon
Nuclear Regulatory Commission
Washington, DC

Carlotta McInteer
Los Alamos Scientific Laboratory
Los Alamos, NM

Michael D. McKay
Los Alamos Scientific Laboratory
Los Alamos, NM

Judith Mahaffey
Pacific Northwest Laboratories
Richland, WA

Sidney Marks
Energy Research and Development Adm.
Washington, DC

Harry Martz
Los Alamos Scientific Laboratory
Los Alamos, NM

M. Mazumdar
Beulah Rand Churchill
Pittsburg, PA

James A. Merrill
Pacific Northwest Laboratories
Richland, WA

F. L. Miller, Jr.
Union Carbide Corporation
Oak Ridge, TN

Roger H. Moore
Nuclear Regulatory Commission
Washington, DC

F. W. Muller
Sandia Laboratories
Albuquerque, NM

Ishwar P. Murarka
Argonne National Laboratory
Argonne, IL

Geronimo Naanep
Los Alamos Scientific Laboratory
Los Alamos, NM

Jan Narud
Los Alamos Scientific Laboratory
Los Alamos, NM

Wesley L. Nicholson
Pacific Northwest Laboratories
Richland, WA

Anthony R. Olsen
Pacific Northwest Laboratories
Richland, WA

C. E. Olson
Sandia Laboratories
Albuquerque, NM

Donald F. Paddleford
Pittsburg, PA

James C. T. Pool
Argonne National Laboratory
Argonne, IL

R. R. Prairie
Sandia Laboratories
Albuquerque, NM

Edward T. Rumble
Science Applications Inc.
Palo Alto, CA

Donald Shepard
Atlantic Richfield Hanford Company
Richland, WA

Nora Smiriga
Lawrence Livermore Laboratory
Livermore, CA

G. P. Steck
Sandia Laboratories
Albuquerque, NM

Don L. Stevens, Jr.
Pacific Northwest Laboratories
Richland, WA

Rodney Strand
Oak Ridge National Laboratory
Oak Ridge, TN

Dante Susco
Los Alamos Scientific Laboratory
Los Alamos, NM

G. L. Tietjen
Los Alamos Scientific Laboratory
Los Alamos, NM

John W. Tukey
Princeton University
Princeton, NJ

V. R. R. Uppuluri
Oak Ridge National Laboratory
Oak Ridge, TN

William Visscher
Los Alamos Scientific Laboratory
Los Alamos, NM

David L. Wallace
University of Chicago
Chicago, IL

Ray Waller
Los Alamos Scientific Laboratory
Los Alamos, NM

Michael S. Waterman
Los Alamos Scientific Laboratory
Los Alamos, NM

Terry A. Watkins
Texas Technological University
Lubbock, TX

David Whiteman
Los Alamos Scientific Laboratory
Los Alamos, NM

William J. Whitty
Los Alamos Scientific Laboratory
Los Alamos, NM

John Wilkinson
Rensselaer Polytechnic Institute
Troy, NY

Michael A. Wincek
Pacific Northwest Laboratories
Richland, WA

R. K. Zeigler
Los Alamos Scientific Laboratory
Los Alamos, NM

Charles Zeis
Rockwell International
Golden, CO

BNWL-1986
UC-32

DISTRIBUTION

No. of Copies

OFFSITE

1 ERDA Chicago Patent Group
Argonne, IL 60439
A. A. Churm

158 ERDA Technical Information Center

1 Division of Reactor Research and Development
ERDA
Washington, DC 20545
Leon Gilford

1 Division of Biomedical and Environmental Research
ERDA
Washington, DC 20545
Sidney Marks

1 Tennessee Valley Authority
Muscle Shroles, AL 35660
Steve Seale

1 Northwestern University
Evanston, IL 60201
Ajit C. Tamhane

3 Argonne National Laboratories
Argonne, IL 60439
Janet Anderson
Marcy L. Hopper
James C.T. Pool
Ishwar P. Murarka

2 Aerojet Nuclear Company
Idaho Falls, ID 83401
Fred J. Balkovetz
Steven J. Bengston

1 Bettis Atomic Power Laboratory
W. Mifflin, PA 15122
William Beggs

1 Battelle
Human Affairs Research Center
Seattle, WA 98105
Carl A. Bennett

No. of Copies

26 Los Alamos Scientific Laboratory
Los Alamos, NM 87545
Kenneth Apt.
R. J. Beckman
Thomas Bement
John Brolley
Larry Bruckner
Bill Buzbee
Katherine Campbell
Jesse Cheadle
F. Andrew Ford
Myrle Horita
David Kahaner
Ronald Lohrding
Victor W. Lowe
Carlotta McInteer
Michael D. McKay
Harry Martz
Geronimo Naanep
Jan Narud
Dante Susco
G. L. Tietjen
William Visscher
Ray Waller
Michael S. Waterman
David Whiteman
William J. Whitty
R. K. Ziegler

9 Sandia Laboratories
Albuquerque, NM 87115
E. E. Ard
J. R. Ellefson
M. A. Griesel
I. J. Hall
Ronald Iman
F. W. Muller
C. E. Olson
R. R. Prairie
G. P. Steck

2 Sandia Laboratories
Livermore, CA 94550
C. W. Childers
C. J. De Carli

No. of Copies

OFFSITE

7 Union Carbide Corp.
Oak Ridge, TN 37830
C. K. Bayne
J. J. Beauchamp
K. O. Bowman
D. A. Gardiner
D. G. Gosslee
W. E. Lever
F. L. Miller, Jr.

2 Princeton Univ.
Princeton, NJ 08540
Peter Bloomfield
John Tukey

5 Texas Tech Univ.
Lubbock, TX 79409
Kamal C. Chanda
W. J. Conover
James M. Davenport
Benjamin S. Duran
Terry A. Watkins

1 Electric Power Research Institute
3412 Hillview Ave.
Palo Alto, CA 94304
Boyer Chu

1 Lawrence Laboratory
Berkeley, CA 94720
Elon Close

1 Inhalation and Toxicology Research
P.O. Box 5890
Albuquerque, NM 87115
Chester R. Crain

1 GE-FBRD
310 De Guigne Dr.
Sunnyvale, CA 94086
Gary L. Crellin

1 Naval Postgrad School
Monterey, CA 93940
Donald P. Gaver

1 Rensselaer Polytechnic Institute
Troy, NY 12181
John W. Wilkinson

No. of Copies

3 Nuclear Regulatory Commission
Washington, DC 20555
R. G. Easterling
Warren Lyon
Roger H. Moore

3 Rockwell International
Golden, CO 80401
Yvonne Ferris
M. Hume
Charles Zeis

1 Stanford Linear Accelerator Center
Stanford, CA 94305
Jerome Friedman

2 Lawrence Livermore Laboratory
Livermore, CA 94550
Perry Gluckman
Nora Smiriga

1 Brookhaven National Laboratory
Upton, L.I., NY 11973
Samuel Kao

1 Westinghouse Electric Corp.
Beulah Rand Churchill
Pittsburgh, PA 15235
M. Mazumdar

1 Box 355
Pittsburgh, PA
Donald F. Paddleford

1 Science Applications Inc.
2680 Hanover Street
Palo Alto, CA 94304
Edward T. Rumble

2 Oak Ridge National Laboratory
Oak Ridge, TN 37830
Rodney Strand
V.R.R. Uppuluri

1 Univ. of Chicago
Chicago, IL 60637
David L. Wallace

1 Oklahoma State Univ.
Stillwater, Oklahoma 74074
Lyle D. Broemeling

No. of
Copies

ONSITE

1 ERDA Richland Operations Office
B. J. Melton

2 Atlantic Richfield Hanford Co.
Daniel E. Griffing
Donald Sheppard

37 Battelle-Northwest
R. L. Buschbom
J. A. Buchanan
L. L. Eberhardt
J. C. Fox
R. O. Gilbert
J. L. Harris (5)
R. L. Hooper
J. W. Johnston
E. L. Kelley, Jr.
V. M. Lee
J. A. Mahaffey
J. A. Merrill
W. L. Nicholson (10)
A. R. Olsen
D. L. Stevens
J. R. Tadlock
P. E. Tucker
C. R. Watson
M. A. Wincek
Technical Information (3)
Technical Publications (1)