JPL Contract 952532

MCR-68-97 (Vol X) Copy No. 8

A Study Program on the Development of a Mathematical Model (s) for Microbial Burden Prediction

Final Report Addendum

Volume X

Phase IX Revisions to Volume VI User's Manual

October 1970

MARTIN MARIETTA CORPORATION

As a result of additional work done in Phase IX (JPL Contract 952532), a new volume, Volume X, has been prepared. Volume X describes changes made in histogram combining and burden prediction and the logic to support these changes. A complete listing of all affected subroutines is included. The results of using the revised MBPM in predicting the microbial burden on the Mariner Mars 69-3 are reported.

(Sheet to be pasted in Volume VI, page ii)

A STUDY PROGRAM ON THE DEVELOPMENT OF MATHEMATICAL MODEL(S) FOR MICROBIAL BURDEN PREDICTION

MCR-68-97

JPL Contract 952532

Volume X

Final Report Addendum

on Phase IX with Revisions

to Volume VI User's Manual

Prepared by:

Lloyd B. Farabee
Program Manager

This work was performed for the Jet Propulsion Laboratory, California Institute of Technology, as sponsored by the National Aeronautics and Space Administration under Contract NAS7-100.

Approved:

George E. Fosdick
Section Chief
Operations Analysis

J. C. Curlander
Manager
Systems Engineering

September 1970

Martin Marietta Corporation
P. O. Box 179
Denver, Colorado 80201

ABSTRACT

The purpose of this study was to refine the histogram method for combining random variables in order to obtain more realistic predictions of the microbial burden on spacecraft. This report describes the basis of histogram operations (adding, multiplying, dividing, subtracting, determining the maximum) to show the nature of the problems involved and the reasons for inaccuracies. A recommended technique is described and a listing of the revised computer program to incorporate this technique into the previously developed Microbial Burden Prediction Model is included. Sample calculations are included to illustrate the operations discussed.

TABLE OF CONTENTS

	PAGE
Abstract	ii
Table of Contents	iii
List of Illustrations	iv
Glossary	1
I. Introduction	2
II. Discussion of Histogram Operations	4
A. General	4
B. Addition of Histograms	8
C. Multiplication of Histograms	16
D. Subtraction of Histograms	18
E. Division of Histograms	21
F. Determining Maximum of Two Histograms	22
G. Input Histograms	23
III. Results of Changes in Histogram Operations	25
A. Effect on Accuracy	25
B. Effect on Run Time	27
C. Comparison of Selected Method with Former Method	27
IV. Conclusions and Recommendations	31
New Technology	33
References	34
Appendixes	36
A. Computer Program Listing of Each Subroutine Affected . .	A-1
B. Computer Program Listing of Control Cards for Detailed Checking	B-1
C. Sample Calculations	C-1

LIST OF ILLUSTRATIONS

Figure 1. Operating on Histograms with Constants

Figure 2. Comparison of Phase VIII MBPM and Swab Sample Predictions of MM 69-3

Figure 3. Revised MBPM Predictions of MM 69-3

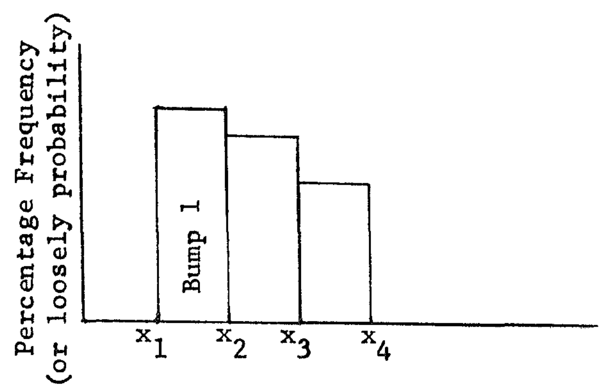
GLOSSARY

1. MBPM: Microbial Burden Prediction Model

2. Class (cell): A convenient division of the range of a variate.

3. Class Limits or Boundaries: The variate values which determine the upper and lower limits of a class.

4. Histogram: A representation of the discrete probability density function (pdf) of a random variable, x. For the purposes of this study, the histograms (pdfs) include but are not limited to, representation of the probability of occurrence versus the number of microorganisms. If the probability density function is continuous, discrete approximations are used. A graphical representation is shown in the figure below.



5. Bump: The portion of the discrete pdf represented by an area between two class limits, (see figure above). The ordinate value, i.e., probability, is assumed constant between the class limit points.

6. Histogram Method: (Also previously called the interval concept.) A numerical technique for performing arithmetic operations of addition, multiplication, division, and subtraction on histograms (pdfs) that are not necessarily from identical underlying distributions.

I. INTRODUCTION

The object of this report is to serve as an aid in understanding the problems involved in performing mathematical operations upon random variables expressed as histograms and as documentation of more satisfactory methods of performing these operations. In order to gain the proper perspective of these problems, it is well to review the methods employed prior to this phase and the reasons, where known and pertinent, for using these methods.

Early in the development of the mathematical model for microbial burden prediction (Ref. 1, pp. 1-4) it was necessary to choose a method for dealing with several random variables in the model. Several methods were available such as closed form solutions, monte carloing, or expected values. With a desire to obtain a complete view of the distributions of the resultant variables with a minimum computer run time, the decision was made to represent these variables as discrete probability density functions (pdf) or histograms (see glossary). The method proposed by JPL to combine these variables was the histogram method. Since the details of the histogram manipulation had not been fully developed, some of the difficulties and limitations of using histograms were not completely understood. This understanding has improved during the preceding phases of this program.

The existence of these problems came to light most dramatically in the tendency, after a large number of successive mathematical operations, of the distributions to concentrate in the highest valued histogram interval (Ref. 4, pp. 35-36) and of the final mean value to be markedly different from the value obtained by performing the same operations on the means of the variables involved (Ref. 4, pp. 36-37). The former was remedied by redefining the output intervals and the latter was resolved by applying the necessary adjustment to the entire distribution by the application of an adjustment factor. This brought the mean back to the desired point and preserved the shape of the distribution, but adjusted the end points of the distribution by the same factor.

The primary objective of Phase IX is to improve the techniques of histogram combining so that this adjustment factor is no longer necessary, or, if this is not possible, to develop a better method of adjustment. Other objectives are to eliminate the subtraction of histograms and to use the revised program to make a burden prediction of the Mariner Mars 69-3, the spacecraft that was used in Phase VIII. Using the MM 69-3 permits an evaluation of the impact of the program changes.

II. DISCUSSION OF HISTOGRAM OPERATIONS

A. General

A histogram is a method of representing a probability density function (either discrete or continuous) by grouping the portions of the distribution together into (usually equal) intervals, with the areas under the ordinate between class limits being proportional to the percentage frequency of occurring between the class limits. For ease in representing these areas graphically, the spacing between class limits is usually a constant (Ref. 11, pp. 5-8). It is not necessary to have equal spacing between class limits when treating the histogram areas mathematically, as in the Microbial Burden Prediction Model. Since the probability between adjacent class limits is represented by a constant, it is inferred that the distribution between adjacent class limits is uniform, although the data upon which the histogram is based seldom satisfy this inference. The theory of histogram combining is based upon the assumption of a uniform distribution between adjacent class limits. Hereafter, in this report, the area between adjacent class limits will be referred to as a "bump".

The equations used in the Microbial Burden Prediction Model (MBPM) which involve mathematical operations on random variables are as follows: (underscored parameters are random variables which are represented as histograms in the computer programs.)

Fallout - - - - - - - - - $\underline{B}' = Be^{-\underline{t}/\underline{v}} + AvR(1 - e^{-\underline{t}/\underline{v}})$ (1)

where $R = f_2 g(\underline{c} + \lambda\ \underline{Q})$ (2)

Contact - - - - - - - - - $\underline{B}' = \underline{B}\left(1 - \frac{a\underline{S}_2}{2A}\right) + \frac{a\underline{S}_1\underline{b}_t}{2}$ (3)

Decontamination - - - - $\underline{B}' = \underline{B}(1 - \underline{k}_j)$ (4)

Burden Difference - - - $B' = \underline{B}_{n+1} - \underline{B}_n$ (5)

where B' is the resulting burden (organisms),

B is the initial burden (organisms),

$e = 2.71828 \ldots$,

t is the activity time (hours),

v is the "average lifetime" (hours),

A is the surface area (sq ft),

R is the fallout rate $(\frac{\text{organisms}}{\text{sq ft-hr}})$,

f_2 is the fallout velocity (ft/hr),

g is the surface retention factor for fallout (dimensionless),

c is the environmental airborne contamination (organisms/cu ft),

Q is the personnel airborne concentration (organisms/cu ft),

λ is a distance reduction factor (1/ft),

d is the distance from the worker to the surface (ft),

a is the area contacted (sq ft),

S_2 is the hand or tool retention factor for contact (dimensionless),

S_1 is the hardware retention factor for contact (dimensionless),

b_t is the contamination on hand or tool (organisms/sq ft),

k_j is the dieoff factor (dimensionless),

B_{n+1} is the burden after n+1 operations

B_n is the burden after n operations

It can be seen by examining these five equations that the following operations involving histograms are included:

(1) Multiplying a histogram by a constant,

(2) Subtracting a histogram from a constant,

(3) Dividing a histogram by a constant,

(4) Exponentiating to a negative histogram power,

(5) Adding one histogram to another histogram,

(6) Multiplying one histogram by another histogram,

(7) Dividing one histogram by another histogram, and

(8) Subtracting one histogram from another histogram.

There is also a ninth operation, determining the maximum of two histograms, which is not apparent from the above equations.

The first four operations are illustrated in Figure 1. Figure 1 (a) shows a histogram, H, where

$P(x_1 \leq H < x_2) = 0.1$

$P(x_2 \leq H < x_3) = 0.3$

$P(x_3 \leq H < x_4) = 0.4$

$P(x_4 \leq H < x_5) = 0.2$

P is the probability

x_1, x_2, x_3, x_4 and x_5 are the class limits.

Figures 1 (b) through (e) show multiplying the histogram (pdf) by a constant c, subtracting from a constant (not necessarily the same) c, dividing by a constant c, and a constant c with a negative histogram exponent, respectively. The area of each interval is proportional to the product of the probability per unit interval (P/U), times the unit length of the abscissa between class limits. Unit interval lengths are constant (equal intervals) for (a), (b), (c), and (d). The unit interval for (e) is $c^{-x_4} - c^{-x_5}$. The value of the cumulative area of each histogram is unity and the abscissa is increasing value from left to right. No error is known to be introduced by these operations which consist of applying the operator to the abscissa without affecting the ordinate (probabilities). The mean value of the histogram is calculated by summing the area-abscissa products. For example, the mean value of the histogram in Figure 1 (a) is

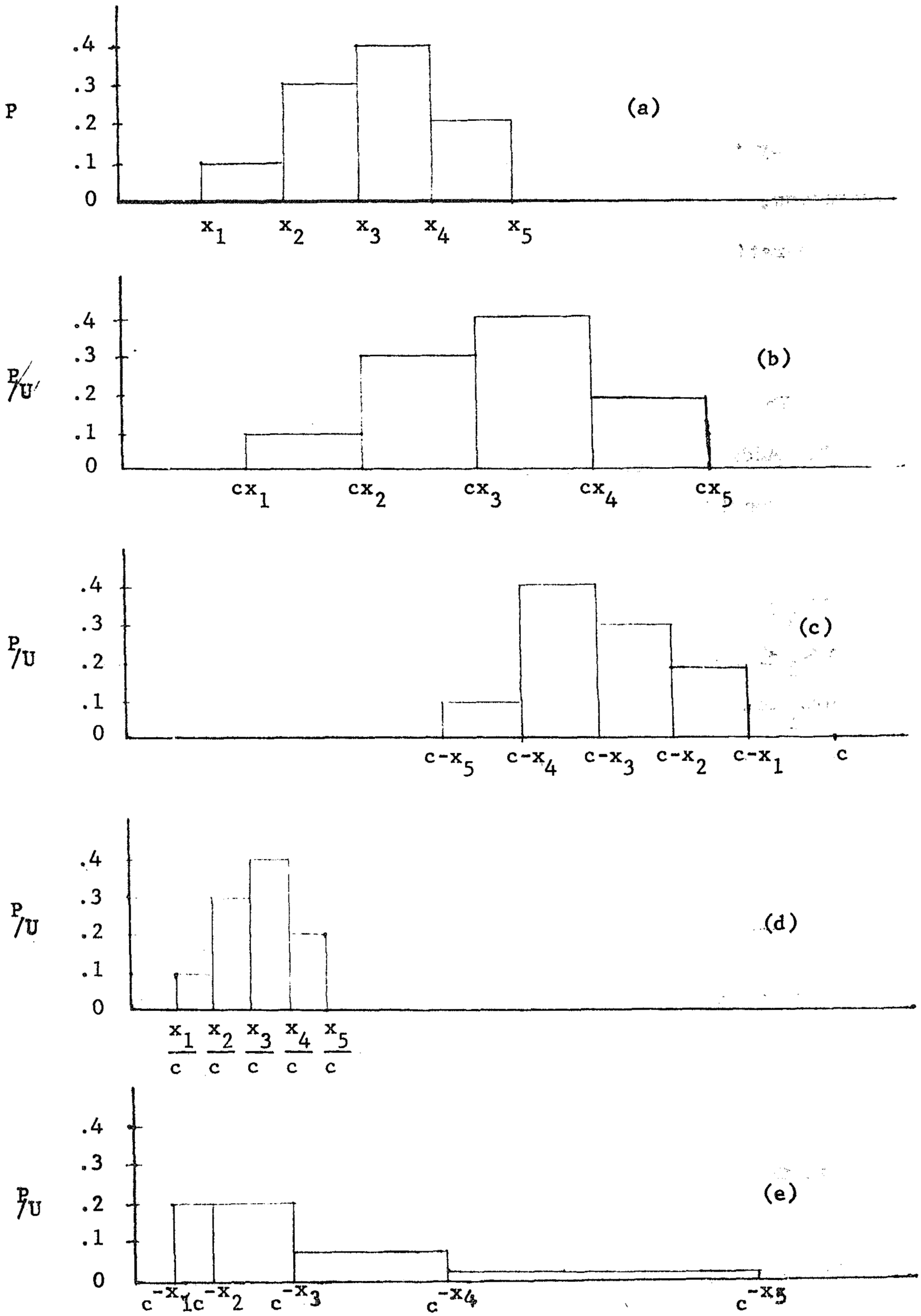


Figure 1. Operating on Histograms with Constants

$$H = 0.1\left(\frac{x_2+x_1}{2}\right) + 0.3\left(\frac{x_3+x_2}{2}\right) + 0.4\left(\frac{x_4+x_3}{2}\right) + 0.2\left(\frac{x_5+x_4}{2}\right) \qquad (6)$$

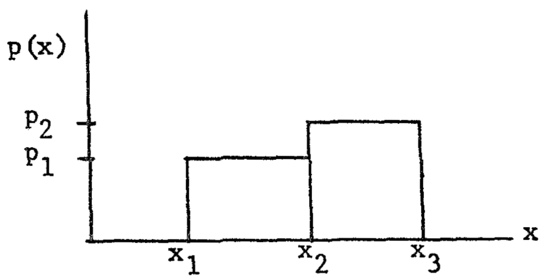
The ninth operation, determining the maximum of two histograms, contains no known errors, and has not been a subject of study in the work covered by this report. Reference 2, pp. 24-27 gives a discussion of this operation.

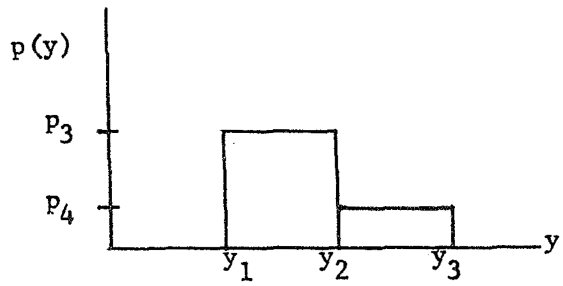
Operations five through eight, which deal with one histogram operating on another histogram, are the ones which have been troublesome in the past and are the primary subjects of investigation in the current phase of the study. These are each discussed in detail in the sections which follow.

B. Addition of Histograms

Consider two independent random variables, x and y, whose distributions are represented by histograms such that

$$P(x_1 \leq x < x_2) = p_1$$
$$P(x_2 \leq x < x_3) = p_2$$
$$P(y_1 \leq y < y_2) = p_3$$
$$P(y_2 \leq y < y_3) = p_4$$

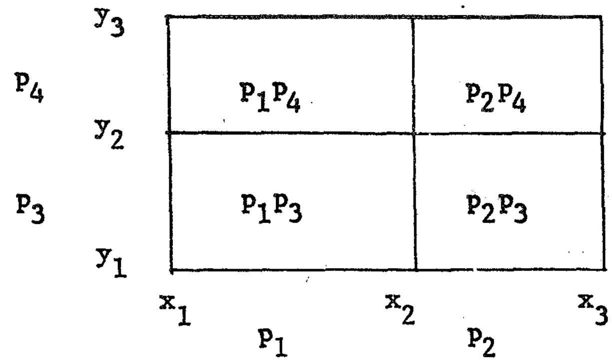




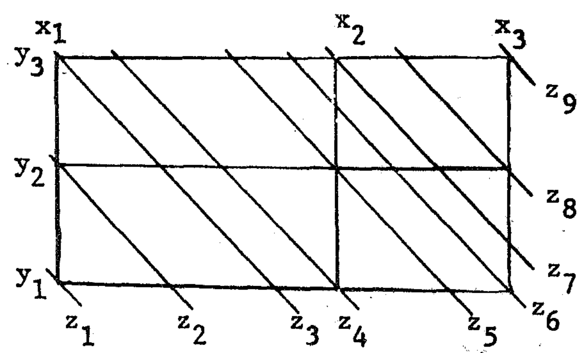
The joint probabilities assuming stochastic independence are

$$P(x_1 \leq x < x_2,\ y_1 \leq y < y_2) = p_1p_3$$
$$P(x_1 \leq x < x_2,\ y_2 \leq y < y_3) = p_1p_4$$
$$P(x_2 \leq x < x_3,\ y_1 \leq y < y_2) = p_2p_3$$
$$P(x_2 \leq x < x_3,\ y_2 \leq y < y_3) = p_2p_4$$

which can be shown diagrammatically as

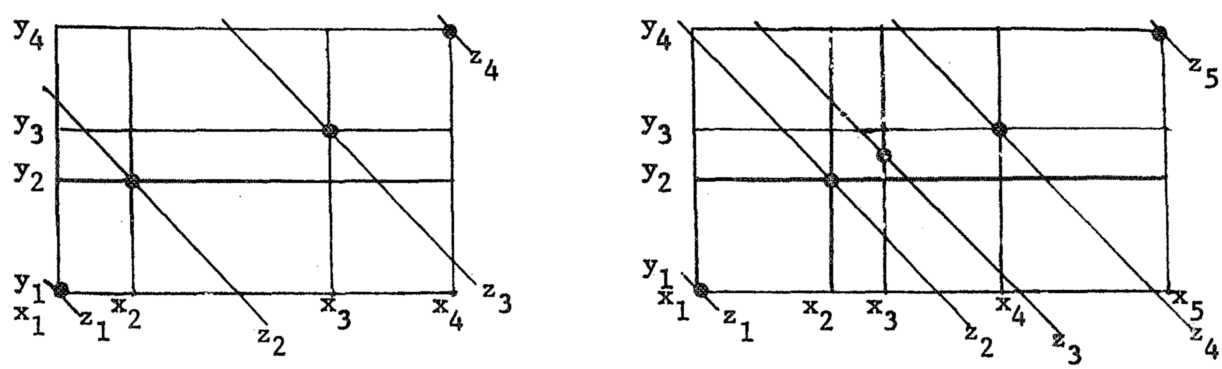


Now, since the distances along the axes are proportional to the values of x and y (and not their probabilities), values of constant z, where z is the sum of the two random variables, $z = x + y$, can be shown as straight lines. Any values of z can be selected. The values in the diagram are the values of z where discontinuities exist because of a change in probability, and, therefore, represent class limits on the z histogram.



Considering the p_1p_3 block as an example, it can be seen that if x is uniformly distributed between x_1 and x_2 and if y is uniformly distributed between y_1 and y_2, then the probability p_1p_3 is uniformly distributed throughout the area, and that the probability of a portion of the area is directly proportional to the portion of the area enclosed. For example, if 1/4 the area of p_1p_3 were between z_1 and z_2, then the probability of z being between z_1 and z_2 would be $\frac{p_1p_3}{4}$. When any z area of interest crosses more than one joint probability segment, as in the case of z between z_2 and z_3, the contributions from the individual segments must be added. This is illustrated in Appendix C.1.

It can be seen from the diagram above that the two-bump by two-bump addition shown results in a z histogram of eight bumps, which is the maximum for two-bump by two-bump. In the general case, where x has n - 1 bumps and y has m - 1 bumps the resulting number of bumps will be nm - 1. It is possible to have fewer than this number if a single z value goes through more than one change of probability point. Unless restricted, this progression in the number of bumps will continue with subsequent operations. In the MBPM, the number of bumps of z is made equal to the larger of the number of bumps on the input histograms by choosing the z values on the intersections as shown below. (Ref. 2, pp. 17-18). The first points chosen are the lower left (x_1y_1) and the upper right (x_4y_4) in the left-hand figure. Other points are chosen by alternately moving up and right from x_1y_1 and moving down and left from x_4y_4 until the point is duplicated. In the case where $n \neq m$ (right-hand figure) when the points x_3y_2 and x_3y_3 were selected, the point half-way between would be the value on x_3, all other points being chosen as described above.



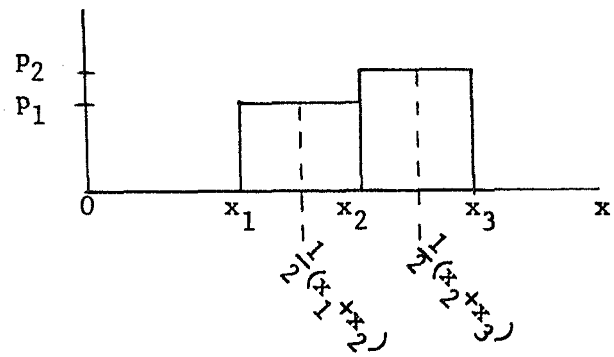
In actual practice, the probabilities are calculated only after the number of bumps is reduced. In the MBPM, prior to Phase IX, an approximation was used in calculating the areas in order to keep computer run time to a minimum. Rather than calculate the exact values of the areas, the areas between z values were taken as proportional to the differences between z values.

Although this approximation seems to introduce only very small errors, even small errors tend to create progressively larger errors after a number of operations. (See Table 1, p. 26). These and other errors were compensated for by applying a correction factor equal to the resultant error in mean value (e.g., -0.05) to all the range values (class limits) of the histogram, while holding the probabilities constant. This preserves the shape of the distribution by sliding the histogram over to where the correct mean value is maintained. The correct mean value is taken to be the sum of the means of x and y, since the mean of the sums should equal the sum of the means. (Ref. 13, p. 208).

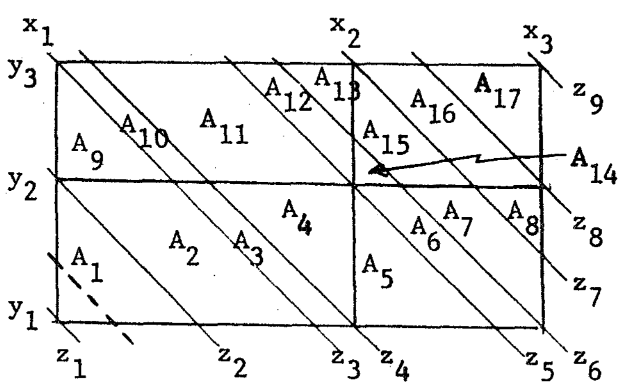
There is, of course, no problem in computing the actual areas rather than the approximate areas, other than the increased run time, and this change has been incorporated into the revised program.

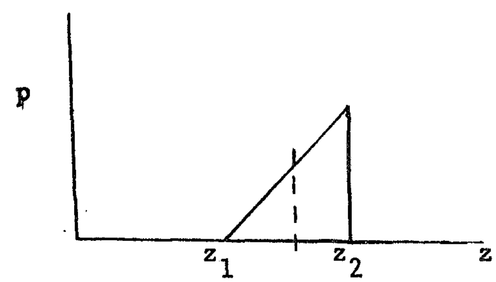
As can be seen from the sample calculations in Appendix C.4, even though the area calculations are correct, the mean is not necessarily correct. An explanation of the method of calculating the mean value of a histogram is helpful in understanding the nature of the occurrence of incorrect mean with correct areas.

The mean of a histogram can be taken as the sum of the means of the individual bumps. The mean of a bump, which is a uniform distribution, is the mean of the class limits. The mean of the variable x in the figure below would be $\bar{x} = \frac{1}{2}(x_1 + x_2)p_1 + \frac{1}{2}(x_2 + x_3)p_2$. This can be seen to be analogous to thinking of the bumps as masses of unit thickness and the mean as the location of the center of gravity. We then determine the physical moment (first statistical moment) about the axis $x = 0$, which is equal to the c.g., or mean, since $p_1 + p_2 = 1$.



If the smaller values of z in the addition diagram are examined, the nature of the problem becomes apparent.

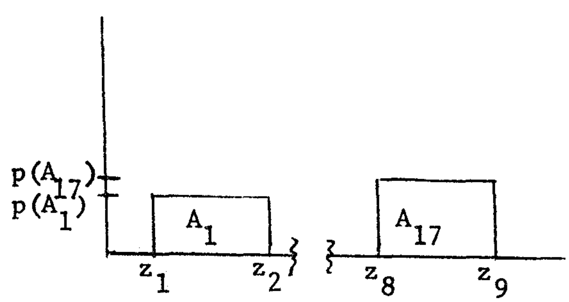




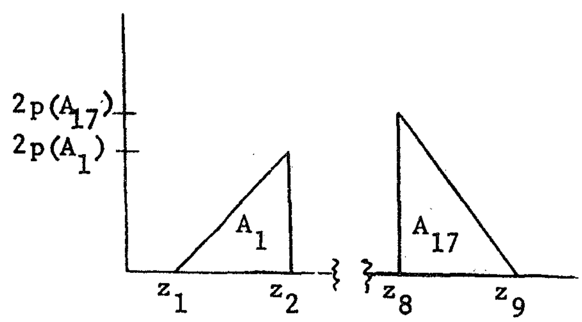
As the dotted line moves in incremental steps from z_1 to z_2 in the left-hand figure the area is not a constant for the steps, but increases linearly. Instead of the distribution between z_1 and z_2 being shown as a rectangle in the right-hand figure, a triangle has been produced.

Since histograms have, by definition, rectangular bumps, we have not produced a histogram bump. In other words, the sum of two histograms is not a histogram. If this is true, the best that can be expected when the sum of two histograms is represented by a histogram is that the resultant will be a good approximation.

By the same logic, a triangle will also be produced in the area between z_8 and z_9. If these areas are represented in the z histogram as rectangular bumps with the limits unchanged, the probabilities of these bumps and the limits of the bumps are correct, but their contribution to the histogram mean is incorrect, due to the shifting of the true mean.

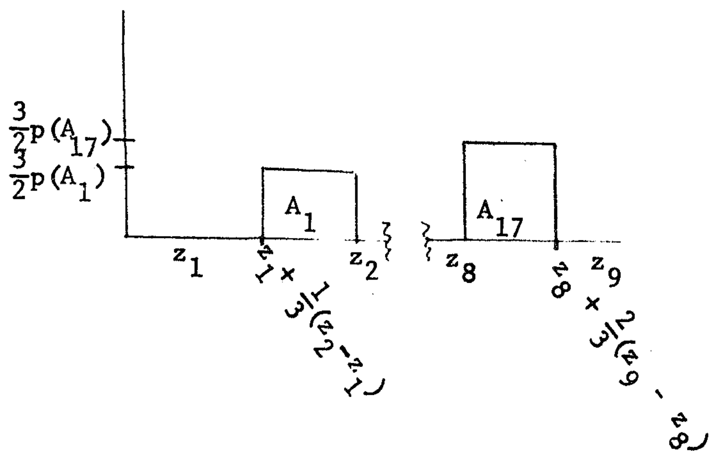


Histogram of
End Bumps



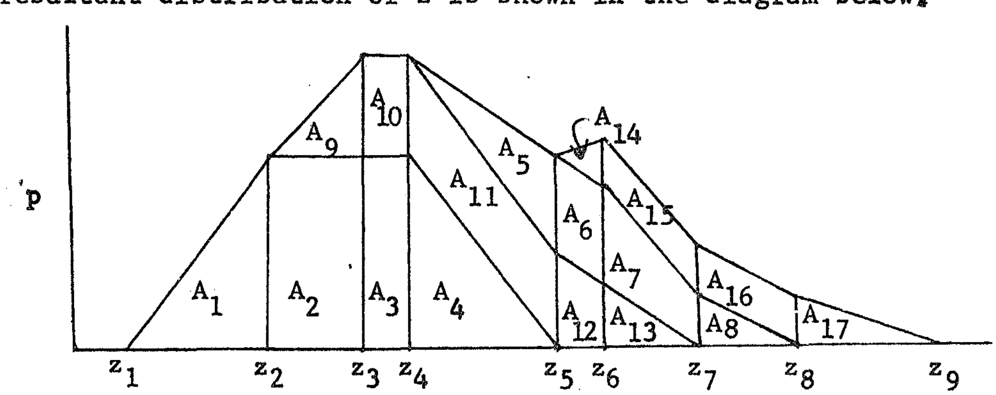
Correct Distribution of
End Bumps

Considering the base of the triangle A_1 (on the joint probability diagram) as z_2 and the base of the triangle A_{17} as z_8, the c.g. of A_1 is at $z_1 + \frac{2}{3}(z_2 - z_1)$ and the c.g. of A_{17} is at $z_8 + \frac{1}{3}(z_9 - z_8)$. The histogram has the c.g. of A_1 at $z_1 + \frac{1}{2}(z_2 - z_1)$ and the c.g. of A_{17} at $z_8 + \frac{1}{2}(z_9 - z_8)$. Obviously, the triangle cannot be replaced by the rectangle and still preserve the correct area, limits, and c.g. The best approximation seems to be to preserve the area and c.g. (which contribute to the histogram mean, the best point of reference) and to make any necessary adjustments on the limits. The only points which can be adjusted are the end values (z_1 and z_9) in this case, as z_2 and z_8 are common to the adjacent bumps. The area affected is also minimal, as the areas are moved from the points z_1 and z_9. The adjusted end points would then be as shown in the diagram below.



An advantage of this method is to counteract the tendency for the one points of the histogram to spread out after repeated operations with an infinitesimal probability of being in this spread portion of the distribution.

An extension of this approach into the other bumps of the eight-bump histogram shows that triangular areas produce triangular bumps and parrallelogram areas produce parallelogram bumps. Other shapes (trapezoids) can be handled as combinations of triangles and parallelograms. The resultant distribution of z is shown in the diagram below.



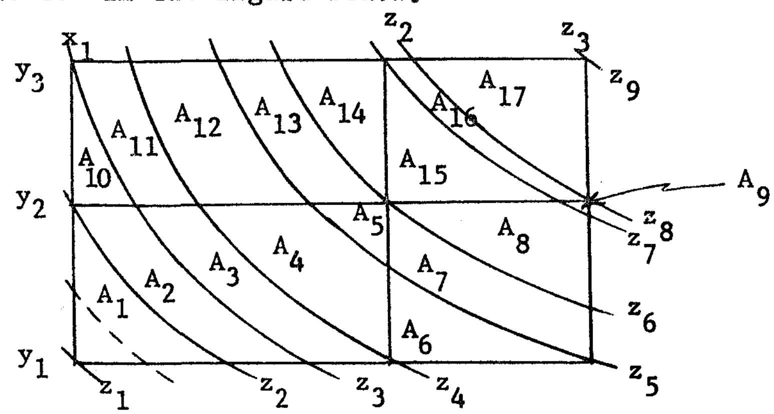
This approach helps to point out another source of error. The sum of the means can be used as a check point against the mean of the sums, but it can be seen (Appendix C.4) that these values can be in agreement and the individual bumps in error. For example, if after reducing the number of bumps, A_1 and A_4 are in the same bump, their equal and opposite errors will balance out, and the bump will be correct, but if z_4 were selected as the dividing line, between bumps, A_1 and A_4 would be in different bumps and each bump would be in error. They would not contribute to error in the histogram mean as long as they were correctly handled (as individual bumps with properly located c.g.) but would contribute when lumped with other values into a single-valued bump. The items that contribute to error in the figure above (where z_5 is the dividing line between bumps for the resultant two-bump histogram) are A_9 and the A_{12}-A_{13} combination, and A_5 and the A_7 -A_8 combination. (See also lower left-hand figure on p. 12.)

It should be apparent that it is impossible to represent the bumps of this figure by rectangular areas (regardless of how the bumps are combined in reducing the number) that preserve the areas, limits and means. Since the areas (probabilities) must total unity and the means are used in the only correction of accuracy, any adjustments should be done to the limits, or dividing lines between bumps. If this is done while preserving the location of the end bumps, the effect will be to skew the z distribution enough to make $\bar{z} = \bar{x} + \bar{y}$. The details of this skewing are illustrated in Appendix C.11.

This method gives a better approximation of the z distribution than the former method (Ref. 2, pp. 15-27) since it preserves the mean value, maintains reasonable end points and distorts the shape as little as possible. (See Appendix C.10 and 11.)

C. Multiplication of Histograms

Consider the same two variables, x and y, which were defined in Section II.B. In finding the distribution of the product $z = (x)(y)$, the joint probabilities of x and y are the same as in addition. On a diagram similar to that used for addition, values of constant z will be hyperbolas as indicated in the figure below.



In general, the philosophy of area being proportional to probability and the manner of reducing the number of bumps after multiplication is the same as in addition. The principal difference lies in the relative difficulty of calculating areas enclosed by hyperbolas as compared to straight lines. This difference suggests an approach to use - to approximate the hyperbolas with straight lines within an area of uniform probability.

In the MBPM prior to Phase IX the same approximation was used for calculating the areas in multiplication that was used in addition, i.e., the areas between z values were taken as proportional to the differences between z values. As can be seen in the sample calculations in Appendix C, the errors introduced by this approximation are larger, in general, when used in multiplication than when used in addition. For example, a comparison of the addition in Appendix C.4 with the multiplication of the same variables in Appendix C.9 shows an error in addition of

$\frac{4.2 - 4.125}{4.2} = 1.79\%$ and an error in multiplication of

$\frac{4.96 - 4.2}{4.2} = 18.1\%.$

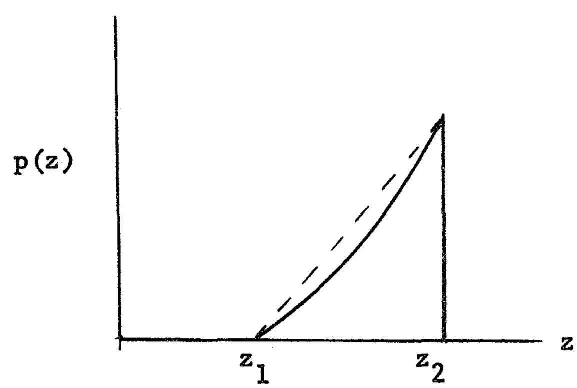
The straight-line method, which has been incorporated into the revised program for multiplication as well as for addition, decreases the error of multiplication but not to the extent that it does for addition, since it eliminates the error of area calculation in addition, but only decreases the error of area calculation in multiplication. The computer run time has been increased in the same manner as in the case of addition.

As the dotted line in the preceding figure moves across A_1 with z increasing, we see that the area is a more complex function of z than it was in addition:

$$A_1 = \int_{x_1}^{z/y} \int_{y_1}^{z/x} dy\, dx = \int_{x_1}^{z/y} \left(\frac{z}{x} - y_1\right) dx \qquad (7)$$

$$= z\left(\ln \frac{z}{y_1} - \ln x_1\right) - y_1\left(\frac{z}{y_1} - x_1\right)$$

When plotted as probability as a function of z, p(z) takes the form shown below:

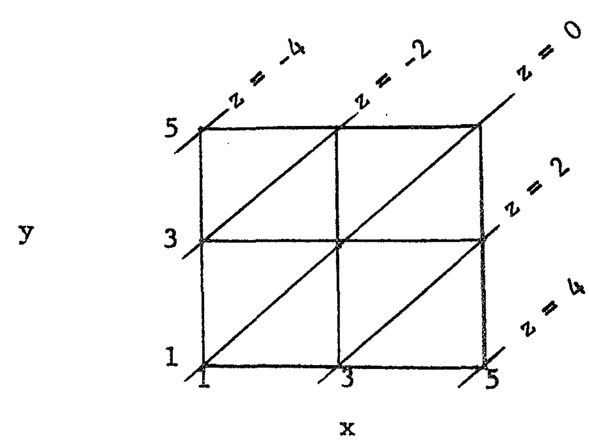


which can be approximated by the straight (dashed) line. If the straight line is accepted as a reasonable approximation, then the method of adjusting the end bumps that was described in Section II.B. can be used with multiplication of histograms.

The fact that the product of the means is equal to the mean of the products of two independent random variables, (Ref. 13, pp. 208-209) provides a means of checking at least one value of the z histogram. Errors in multiplication tend to be greater than in addition for the reasons already mentioned and also because there is not the symmetry of equal triangles that tends to provide equal and opposite errors which balance out, and errors tend to be positive since the dashed line in the above figure includes more than the correct area. Otherwise the same general philosophy of approximating the product of two variables can be used as was used for approximating the sum of two variables. The same (although probably larger) skewing effect will be exerted while retaining the proper mean and realistic end points.

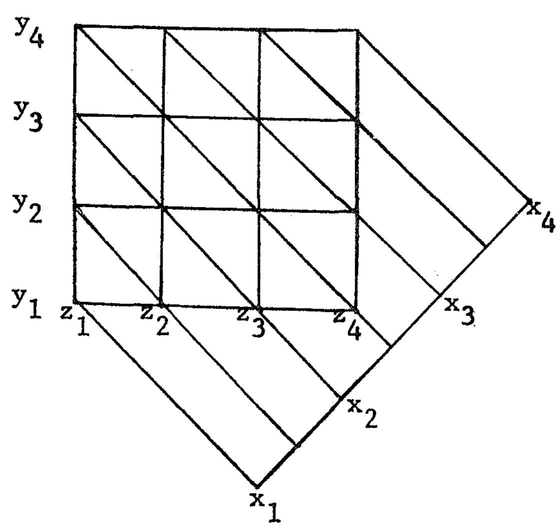
D. Subtraction of Histograms

It would seem that subtraction of histograms would be a quite simple operation, comparable to addition. Although conceptually simple, such a process contains problems of implementation. For example, in the diagram below, treated in a manner similar to addition and multiplication, negative values of $z = x - y$ are possible.



Negative values of burden are, of course, meaningless, and a method of addition was adopted.

In the MBPM the distribution of z is the quantity which, when added to y (which has some known distribution) will give the value x with a given distribution. A discussion of the method of subtraction as used in the MBPM is given in Reference 2, pp. 21-24. It may be summarized in the following diagram where x and y are the known variables, and z is the variable which is added to y to produce x in the equation $x = y + z$ (for $z = x - y$).



There are several problems with using this method, including the following:

(1) With some combinations of x and y it is still possible to get negative values of z;

(2) This method requires dependence of x and y, but y and z seem to retain independence;

(3) Not only must x and y be dependent, but they must be dependent in a manner that will give a mathematically consistent value for z;

(4) In an example given in Reference 2 where x and y are identical, z must equal 0 with probability 1 which is difficult to justify in a real situation, such as the decontamination process.

In view of these problems inherent to histogram subtraction, the places where subtraction is used in the MBPM were examined and found only in the "burden change" calculation and in the decontamination calculation.

The burden change calculation is performed so that a printout might be obtained after any task, but it is not in a direct line with subsequent burden calculations. In other words, the results of this calculation are not used in any other calculations, but are for information only. Since the value of this information is not significant, this calculation can be eliminated from the program with no serious consequence.

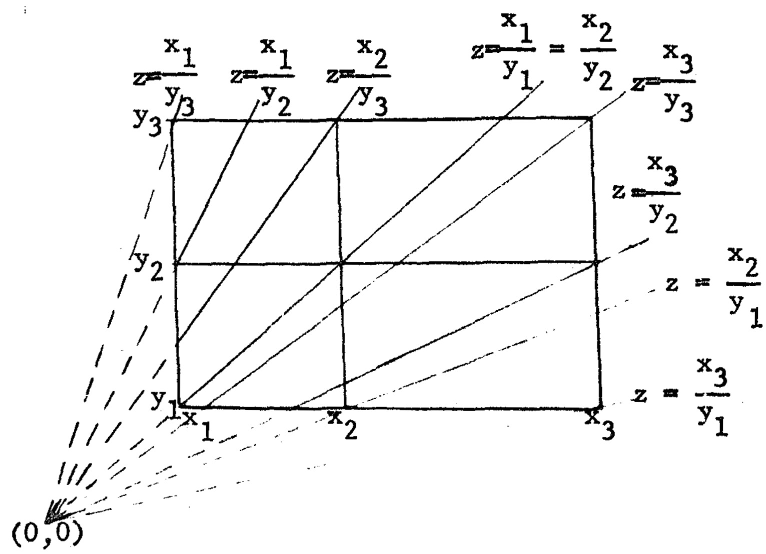
Equation (3), which is used to calculate the contamination caused by contact with hands or tools, was performed in the MBPM as $B' = B - \frac{aS_2B}{2A} + \frac{aS_1b_t}{2}$, which required subtracting one histogram from another. This is being changed to $B' = B\left(1 - \frac{aS_2}{2A}\right) + \frac{aS_1b_t}{2}$ as given in equation (3). The operations will now be subtracting a histogram from a constant (which, as pointed out in Section II.A, presents no known problems), multiplication of two histograms and adding two histograms.

The problems of histogram subtraction are being avoided by eliminating the requirement for this operation from the MBPM. The only subtractions remaining are the subtraction of a histogram from a constant.

In equation (1) $e^{-t/v}$ is always less than one, and in equation (3) $aS_2/2A$ is always less than one. A check has been incorporated to show an error if k_j in equation (4) is input greater than one. As these three values are subtracted from unity, the resultant is always a positive value. The problems of histogram subtraction have thus been avoided without compromising any of the required calculations.

E. Division of Histograms

The division of one histogram by another histogram was performed in the MBPM in a manner similar to addition and multiplication and is illustrated in the diagram below for $z = x/y$.



Division seemed to present no new problems, but when checking or adjusting the resultant a problem arose, since the mean of the quotients is not equal to the quotient of the means ($\bar{z} \neq \bar{x}/\bar{y}$). (Ref. 13, p. 224). The choice, then, was either to develop a new method of acquiring a correction factor, or to eliminate the requirement for division.

The only requirement for histogram division in the MBPM is in the exponent -$\underline{t}/\underline{v}$. The most logical way to make this calculation is to take the reciprocal of $\underline{v}$, ($1/\underline{v}$), which is simply a constant divided by a histogram, and multiply the resultant histogram by the histogram for $\underline{t}$. The mean of $1/\underline{v}$ can be determined in a straightforward manner, and the mean of the products is equal to the product of the means. This has been incorporated into the MBPM, thereby eliminating the need to divide one histogram by another histogram.

F. Determining Maximum of Two Histograms

The method of determining the maximum of two histograms is described in Reference 2, pp. 24-26, but may be summarized by saying that the probability that two independent events will have occurred by some time (t) is the product of the probabilities that each has occurred. For any pair of values x and y the cumulative probability for $z = \max (x,y)$ is the product

$$P_z(z) = P_x(z) \cdot P_y(z)$$

The MBPM uses the method in Reference 2 to determine values for z on a cumulative probability curve and then eliminates the least significant points on the curve to reduce the number of z points to the maximum number of x points or y points.

Since no errors have been detected in this method, no changes have been incorporated into the program. It should be pointed out that this feature of the program has not been exercised on recent MBPM runs, since in the only place this is used in the program (operation times) the resultant operation time has been a known quantity for each operation. It is anticipated that this calculation will be used extensively on future runs, however, and should be retained in the program.

G. Input Histograms

In order to get to the point where histogram combining is necessary, the random variables must be expressed as histograms. The question of how to go from a sample of data to a histogram to be used in the MBPM has been the subject of some criticism. As pointed out previously, the inference of uniform distribution between class limits of a histogram is seldom even approximated by the data upon which the input histograms are based.

As is shown in Appendix C.1 and 2, input data can be represented by more than one histogram for the same sample with widely varying results when combined in histogram operations. For this reason, the selection of class limits can be an important consideration. Questions which frequently arise are, for example, into how many bumps should a histogram based on given data be broken; should equal or unequal class sizes be used; if unequal, should they be logarithmic; how far above the highest data point should the top class go; how far below the lowest data point should the bottom class go? In the past, these questions have been answered on the basis of professional judgment, except that computer memory size limitations restrict the maximum number of bumps per histogram to about ten for the Mariner type of input data.

It has been suggested that some of the input histograms might be expressed appropriately on a logarithmic base. Some of the data do seem to be based on a logarithmic normally distributed base, but the advantage of using this base in preference to a more linear one has not been demonstrated on the MBPM.

The input data used during this phase of the study were essentially that which were reported in Reference 4. These data were based on limited experimental data and considerable engineering judgment and the need for improvement is recognized by the current program at JPL to revise and update all input histograms. Graphical tests were used (Appendix C.12) as a measure of the normality of the input data and it was found that most, if not all, of the input data are approximated by either a normally or logarithmic normally distributed curve. It is felt that further effort in improving the input data is pertinent to increasing the accuracy of the predictions obtained through the use of the MBPM.

Another area that should be reviewed is the validity of the input data. For example, two sources of data (Ref. 4, pp. 9-10) were used to obtain the input histogram for the average microbial lifetime, v. One source consists of eight data points with a mean of 230 hours, while the other source consists of 69 data points with a mean of 30 hours. The combination of these 77 data points which has a mean of 51 hours is used as the basis for the input histogram. The apparent disparity of the source data upon which the input histograms are based needs to be resolved by further study of the validity of the source data which is outside the scope of this contract.

III. RESULTS OF CHANGES IN HISTOGRAM OPERATIONS

A. Effect on Accuracy

The effect of changing the way the computer program performs histogram operations on the overall accuracy of the MBPM calculations has not been defined, but the following statements do apply to individual steps of histogram combining used in the MBPM:

(1) The straight-line method of calculating probability is more accurate than the proportional method in calculating the shape and the mean of products and is more accurate in calculating the shape of sums, but is less accurate in calculating the mean of sums;

(2) The method of fixing end points by preserving the areas and moments of the end bumps tends to prevent the buildup of excessively long "tails" and gives a more realistic picture of the distribution;

(3) The method of skewing by applying a correction factor to all class limits except the end points preserves the range values better than the method of applying the factor to the entire distribution;

(4) Replacement of histogram division by multiplication with the reciprocal of the histogram improves the accuracy since there is no known way to easily correct the result of division to a known mean;

(5) Elimination of histogram subtraction improves the accuracy by removing some non-rigorous procedures from the program and by reordering the sequence of operations.

(6) When skewing will not accomplish the necessary correction, the program uses the correction factor essentially the way it was used prior to Phase IX and the results are neither better nor worse than before the revision. This can become necessary when multiplying one histogram by another histogram several orders of magnitude larger in range values.

A detailed check of histogram combining results was made by using the listing shown in Appendix B. The results below are for the multiplication of M identical normally distributed histograms, each with a mean of 4.0, using the straight-line method of multiplication, without any other corrections or adjustments.

M	No. of Bumps	True Value of Mean	Calculated Value of Mean
2	3	16.0	17.697
2	5	16.0	16.48
2	10	16.0	16.131
5	3	1.02×10^3	3.298×10^3
5	5	1.02×10^3	1.676×10^3
5	10	1.02×10^3	1.173×10^3
10	3	1.05×10^6	8.562×10^7
10	5	1.05×10^6	1.193×10^7
10	10	1.05×10^6	2.552×10^6

TABLE I. EFFECT OF VARYING NUMBERS OF OPERATIONS AND BUMPS

The effect of varying the number of bumps and the effect of repeated multiplications can be seen from the above table. The desirability of using a large number of bumps, and the necessity of correcting the mean is apparent.

B. Effect on Run Time

The effect of the program revisions of histogram operations on computer run time is expected to vary slightly with the type of problem being run and with the computer being used. The table below is for two runs on the CDC 6500 at the Martin Marietta Denver facility. The first run is with the program as used in Phase VIII (the burden difference calculation has been eliminated to give a truer picture of the calculation revisions) and the second run is with the program with all the revisions adopted during Phase IX. The difference in run time is taken to be the total central processor time less the load time. The times are for the first ten days of the Mariner Mars 69-3.

Run No.	Total Time (sec.)	Load Time (sec.)	Run Time (sec.)
1	100.591	29.421	71.170
2	111.073	30.128	80.945

TABLE II. EFFECT OF PROGRAM REVISIONS ON COMPUTER RUN TIME

The increase in run time for this case is approximately 14%. A similar run on the Univac 1108 at JPL increased run time from 836 to 956 sec. or 14%. These are indicative of what might be expected on other runs.

C. Comparison of Selected Method with Former Method

Figure 2 shows the former MBPM predictions compared with swab sample data for the Mariner Mars 69-3. These are the same data that were reported in Reference 9, Figure 1. Figure 3 shows the revised MBPM predictions for the same Mariner Mars 69-3 input data.

There is no significant difference in the mean values. The revised MBPM gives slightly higher values in the vicinity of 120 days, but from 150 to 250 days the means are essentially equal. The 5 and 95 percentile curves tend to be closer to the mean value in the revised MBPM than in the former method. This indicates that the mean value (which is calculated in basically the same way as before) has not been affected, while the revised method of calculation has affected the **shape** of the histograms.

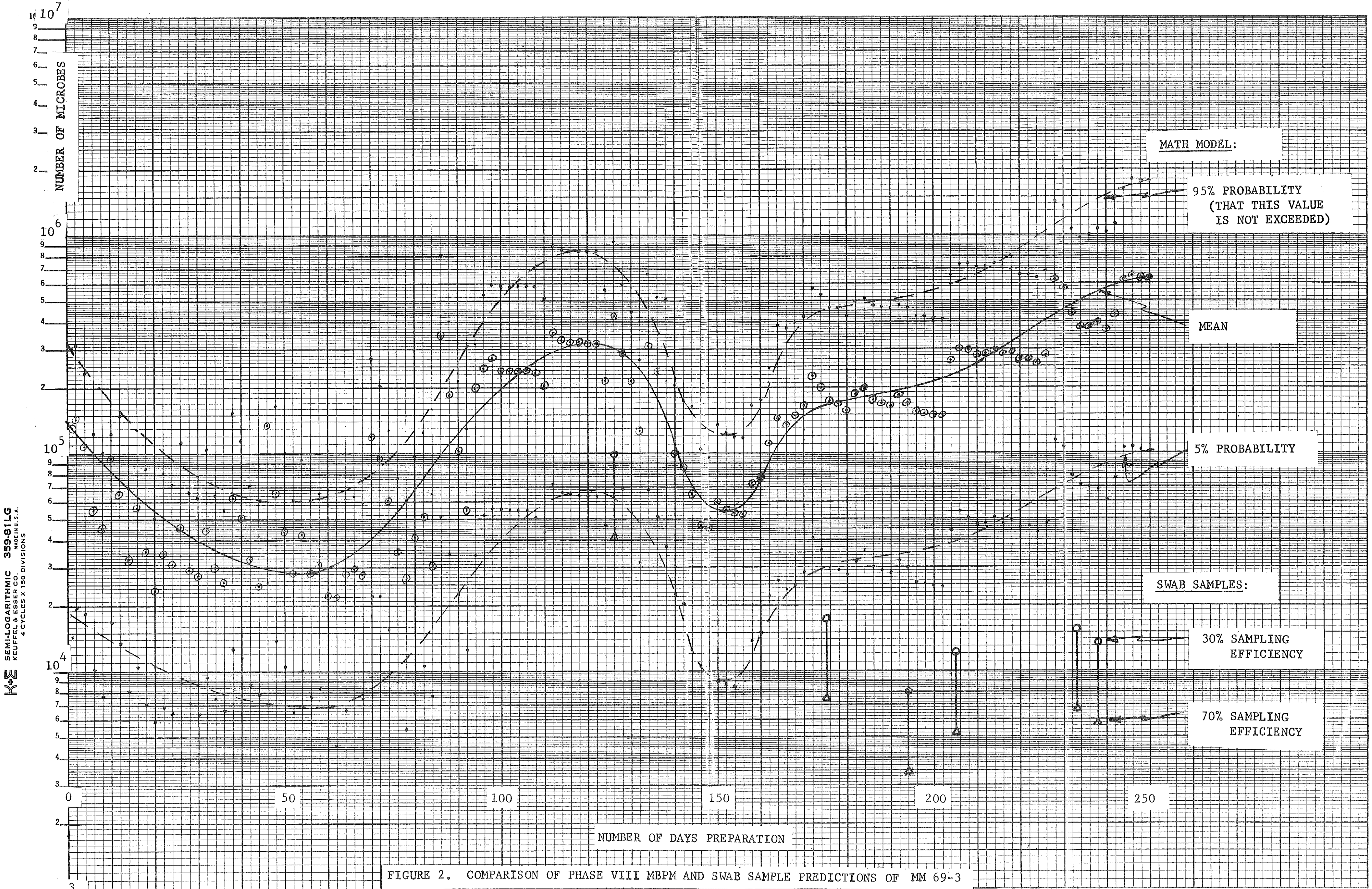


FIGURE 2. COMPARISON OF PHASE VIII MBPM AND SWAB SAMPLE PREDICTIONS OF MM 69-3

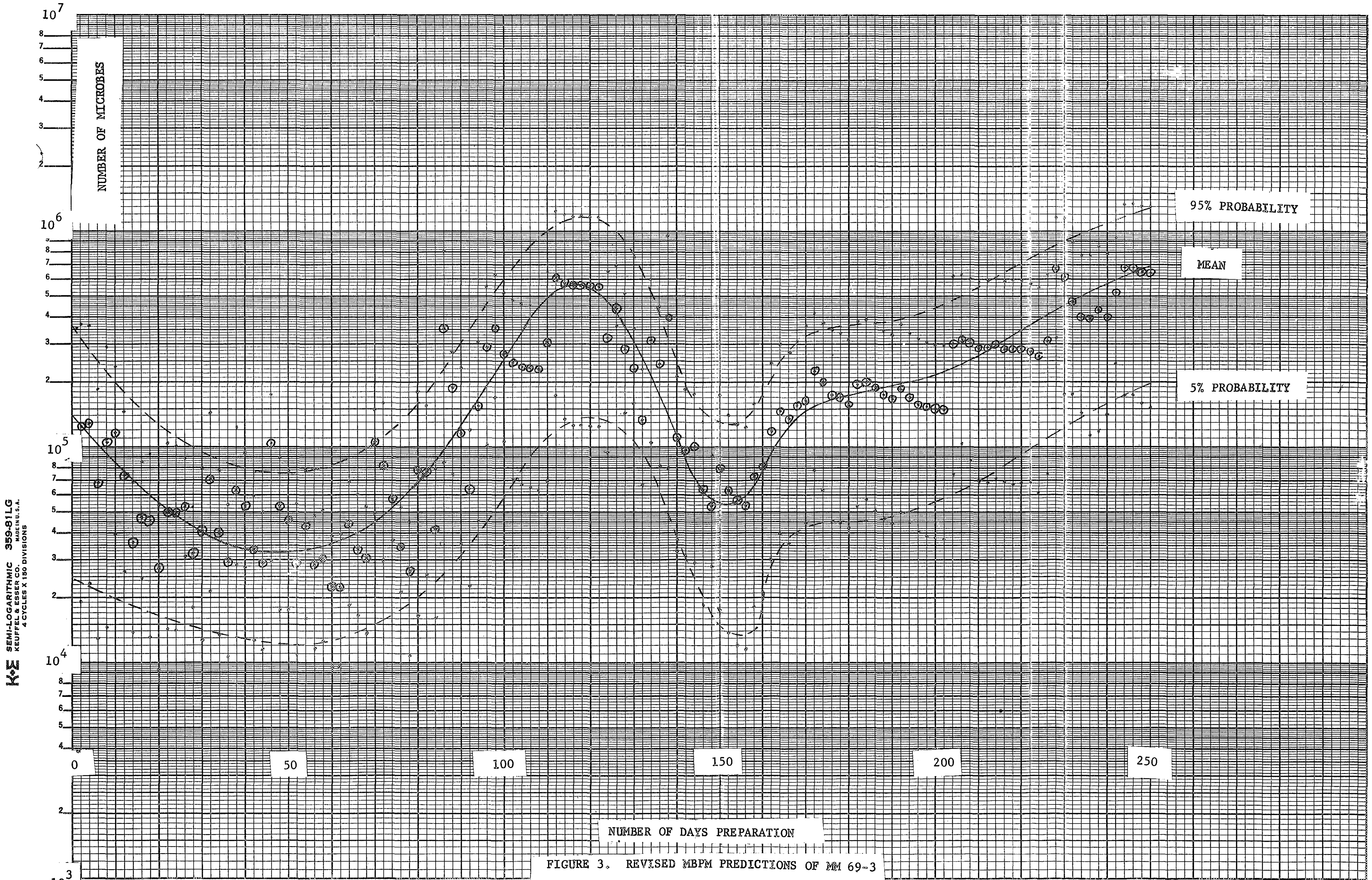


FIGURE 3. REVISED MBPM PREDICTIONS OF MM 69-3

IV. CONCLUSIONS AND RECOMMENDATIONS

A mathematically precise technique for combining histograms into sums, products, difference and quotients, which are also histograms, is not available. The absence of this most desirable way of performing these operations leads to a search for a method which will give the best approximation while preserving the primary characteristics of the distribution as much as is practical. These characteristics have been taken to be mean value, distribution shape, and end point values, although only the mean value can be expressed as a quantitative measure of accuracy.

The program revisions which have been made during this tudy lead to the following conclusions:

(1) Additional insight has been gained into the nature of the histogram combining problem and this insight had led to a method of calculation that maintains the extremities of the distribution without sacrifice of mean value or adversely affecting the shape of the distribution. This should give a more realistic picture of the distributions than the method formerly used.

(2) Since the method of calculating means has not been altered (except for division and subtraction, which have been eliminated) the resultant means are basically unchanged.

(3) Since computer run time did not seem to be seriously increased by the method which incorporated all the improvements developed during the study, all the improvements were incorporated into the revised MBPM.

It has not been within the scope of the present contract to make an exhaustive investigation into histogram combining techniques. If histogram combining is to continue being used in probabilistic studies, more rigorous methods should be investigated. It might be desirable to perform the combining calculations in some more generalized manner and express the resultant as a histogram.

NEW TECHNOLOGY

This study reports the development of revisions to a previously reported mathematical model. The histogram method for combining random variables that was refined in this study has potential applications in other engineering and scientific fields.

REFERENCES

The references (1) through (5), which were issued in December, 1968, and are available through the NASA STAR, form the final report for JPL Contract 952028, A Study Program on the Development of a Mathematical Model(s) for Microbial Burden Prediction:

(1) Volume I, Technical Report

(2) Volume II, User's Manual for the Microbial Burden Prediction Model

(3) Volume III, Appendices

(4) Volume IV, Addendum: Technical Report

(5) Volume V, Addendum: Appendices

The reference (6), which was issued in October 1969 forms the report for the work performed under the original phase of Contract 952532:

(6) Volume VI, Technical Report, Phase VII, Revised User's Manual for the Microbial Burden Prediction Model

The references (7) through (9) report the work performed under the first extension to Contract 952532:

(7) Volume VII, User's Manual for the Input Translator Program

(8) Volume VIII, Addendum to the User's Manual for the Microbial Burden Prediction Model

(9) Volume IX, Final Report Addendum, Report on the Mariner Mars 69-3 Burden Prediction

(10) Donahue, James D., "Products and Quotients of Random Variables Their Applications", Office of Aerospace Research, United States Air Force, ARL 64-115, July 1964.

(11) W. J. Dixon and F. J. Massey, Jr., Introduction to Statistical Analysis, McGraw-Hill.

(12) A. M. Mood and F. A. Graybill, Introduction to the Theory of Statistics, McGraw-Hill.

(13) William Feller, An Introduction to Probability Theory and Its Applications, Wiley.

(14) British Association for the Advancement of Science, Mathematical Tables, Vol. VI, Bessel Functions, Part I, Functions of Orders Zero and Unity, University Press, Cambridge.

Appendix A

COMPUTER PROGRAM LISTING OF EACH SUBROUTINE AFFECTED.

```
      SUBROUTINE BPS
C     BURDEN PREDICTION SUBROUTINE
C     THIS PROGRAM READS FROM TAPE 9
      COMMON    KK, IK, KR, RUN(7), KS, STG(7), KT, TSK(7), DSC(6), ID,
     .          NE, AES(4), IEC(10), IET(10), AEC(10), AET(10), AED(10),
     .          AEF(10,4), NO, IOT(35), IOQ(35), IOC(35), NP, DAC(120,2),
     .          IU(120), IAB(120,6), AAG(120), AAS(120), AAT(120),
     .          NX(572), DR(572,11), XR(572,11), KO, NS, IR(5), AR(4),
     .          APA(2), NT, JT(100), XMT(100), XVT(100), L1, L2, L3, L4,
     .          KTS, ITE, K3, DUM(2664)
      DIMENSION XX(7),IX(6)
    3 FORMAT(34H1MICROBIAL BURDEN PREDICTION MODEL/4H RUN,I5,2H, ,7A6)
      REWIND 9
      REWIND 11
      NT=0
  100 READ (9) L,KK,IK,I1,I2,I5,I6,XX
C     KK=1 INDICATES A NEW RUN, STAGE, AND TASK
C     KK=2 INDICATES A NEW STAGE AND TASK
C     KK=3 INDICATES A NEW TASK
C     KK=4 CALLS FOR A BURDEN DIFFERENCE DETERMINATION
C     KK=5 CALLS FOR ZONE BURDEN WRITEOUT
C        IK=0 CALLS FOR ZONE DEFINITION DATA INPUTS
C        IK=1 INDICATES ZONES CORRESPOND TO PARTS
C     KK=6 CALLS FOR A RESTART AND REQUIRES A BINARY DECK INPUT
      IF(KK.NE.6)GO TO 102
      READ(2) KR,(RUN(J),J=1,7),KS,(STG(J),J=1,7),NE,(AES(J),J=1,4),        A
     . (IEC(J),IET(J),AEC(J),AET(J),AED(J),(AEF(J,K),K=1,4),J=1,10),        G
     . NO,(IOT(J),IOQ(J),IOC(J),J=1,35),NP,((DAC(J,K),K=1,2),IU(J),         G
     . (IAB(J,K),K=1,6),AAG(J),AAS(J),AAT(J),J=1,100),KO,NT,                G
     . (JT(J),XMT(J),XVT(J),J=1,100),KTS,ID,ITE,K1,K2,                      G
     . (NX(I),(DR(I,J),XR(I,J),J=1,11),I=1,572)                             G
      GO TO 100
  102 IF(KK-3) 103,130,109
  103 IF(NT.EQ.0)GO TO 106
C
C     WRITE STAGE SUMMARY FOR PRECEEDING STAGE
      WRITE(6,3)KR,(RUN(J),J=1,7)
      WRITE(6,4)KS,(STG(J),J=1,7)
    4 FORMAT(6H STAGE,I3,2H, ,7A6/15H STAGE SUMMARY-//
     .3X,4HTASK,7X,4HMEAN,6X,6HFINISH/13X,6HBURDEN,6X,4HTIME/)
      DO 104 I=1,NT
  104 WRITE(6,5)JT(I),XMT(I),XVT(I)
    5 FORMAT(I7,E13.3,F10.3)
  106 IF(KK.GT.0) GO TO 109
      IF(K2.EQ.3) GO TO 108
C
      WRITE(3) KR,(RUN(J),J=1,7),KS,(STG(J),J=1,7),NE,(AES(J),J=1,4),       G
     . (IEC(J),IET(J),AEC(J),AET(J),AED(J),(AEF(J,K),K=1,4),J=1,10),        G
     . NO,(IOT(J),IOQ(J),IOC(J),J=1,35),NP,((DAC(J,K),K=1,2),IU(J),         G
     . (IAB(J,K),K=1,6),AAG(J),AAS(J),AAT(J),J=1,100),KO,NT,                G
     . (JT(J),XMT(J),XVT(J),J=1,100),KTS,ID,ITE,K1,K2,                      G
     . (NX(I),(DR(I,J),XR(I,J),J=1,11),I=1,572)                             G
  108 M=0
C
      WRITE(11) M,KR,KS,M,M,(DSC(J),J=1,4),M,M,(RUN(J),J=1,7),(STG(J),
     .J=1,7),(JT(J),J=1,18)
      RETURN
  109 GO TO (110,120,130,140,150),KK
C
C     READ RUN NUMBER AND DESCRIPTION
  110 READ (9) L,KR,(IX(J),J=1,5),RUN
      WRITE(6,600) KR,(RUN(J),J=1,7)
  600 FORMAT(34H1MICROBIAL BURDEN PREDICTION MODEL//4H RUN,I2,2X,7A6//)
```

```
      ID=0
      MD=0
      K1=0                                                              G
      K2=1
      NE=0
      NO=0
      NP=0
      NT=0
      DO 111 I=1,120
      IU(I)=0
      AAG(I)=0.
      AAS(I)=0.
      AAT(I)=0.
      DO 111 J=1,6                                                      G
  111 IAB(I,J)=0
      DO 112 N=1,572
  112 NX(N)=-1
      KO=508                                                            G
      NX(KO)=1                                                          G
      DR(KO,1)=0.                                                       G
      XR(KO,1)=0.                                                       G
C
C     READ STAGE NUMBER AND DESCRIPTION
  120 READ (9) L,KS,(IX(J),J=1,5),STG
      NT=0
      M=1
C
      WRITE(11) M,KR,KS,M,M,(DSC(J),J=1,4),M,M,(RUN(J),J=1,7),(STG(J),
     .J=1,7),(JT(J),J=1,18)
C
C     READ TASK NUMBER AND DESCRIPTION
C     L4 IS THE INDEX FOR SAVING THE TOTAL MICROBIAL BURDEN
C        IF NEEDED FOR USE IN DETERMINING A BURDEN DIFFERENCE
C        K3 IS BURDEN PRINT FLAG                                        G
  130 READ(9) L,KT,L4,K3,(IX(J),J=1,3),TSK                              G
      WRITE(6,612) KS,STG,KT,TSK
  612 FORMAT(6H1STAGE,I3,2X,7A6/5H TASK,I4,2X,7A6/)
      IF(I1.GT.0)K1=I1
      IF(I2.GT.0)K2=I2
      IF(I5.LE.0) GO TO 138
      IF(ID.GT.0) MD=NX(ID)
      ID=I5
C  RESTORE THE FIRST 20 DISTRIBUTIONS
      DO 137 K=1,20
      J=K+552
      CALL HES(K,J)
  137 CONTINUE
  138 IF(I6.GT.0)ITE=I6
      NT=NT+1
      CALL MBS(MD,K1,K2)
      IF(K2.NE.2) GO TO 100
C
      WRITE(3) KR,(RUN(J),J=1,7),KS,(STG(J),J=1,7),NE,(AES(J),J=1,4),   G
     . (IEC(J),IET(J),AEC(J),AET(J),AED(J),(AEF(J,K),K=1,4),J=1,10),    G
     . NO,(IOT(J),IOO(J),IOC(J),J=1,35),NP,((DAC(J,K),K=1,2),IU(J),     G
     . (IAB(J,K),K=1,6),AAG(J),AAS(J),AAT(J),J=1,100),KO,NT,            G
     . (JT(J),XMT(J),XVT(J),J=1,100),KTS,ID,ITE,K1,K2,                  G
     . (NX(I),(DR(I,J),XR(I,J),J=1,11),I=1,572)                         G
C
C     DETERMINE AND WRITE BURDEN DIFFERENCE
  140 WRITE (6,601)
  601 FORMAT ( // 35H BURDEN DIFFERENCE HAS BEEN DELETED. // )
      GO TO 100
```

```
C
C      DETERMINE AND WRITE ZONE BURDEN DISTRIBUTIONS
 150   WRITE(6,3)KR,(RUN(J),J=1,7)
       WRITE(6,7)KS,KT
    7  FORMAT(43H MICROBIAL BURDEN BY ZONES (FOLLOWING STAGE,I3,6H, TASK,
      .I3,2H)-/)
       IF(IK.EQ.1)GO TO 180
C
C      USE ZONE DEFINITION INPUTS
  155 READ (9) L,IZ,(IX(J),J=1,5),DSC,XX(1)
       IF(IZ.LE.0)GO TO 100
       A=0.
       CALL HES(510,KO)
  160 READ (9) L,IP,(IX(J),J=1,5),AR,(XX(J),J=1,3)
C      IP IS THE PART
C      AR(J) THE FRACTION OF SURFACE J OF PART IP BELONGING TO ZONE IZ
       IF(IP.LE.0)GO TO 170
       DO 165 J=1,4
       IF(AR(J).EQ.0.)GO TO 165
       IF(IAB(IP,J).EQ.0)GO TO 165
       F=AR(J)
       IB=IAB(IP,J)
       CALL HMS(IB,F,509)
       A=A+DR(509,1)
       DR(509,1)=AVF(509)
       CALL HCS(510,509,510,1)
 165   CONTINUE
       GO TO 160
 170   WRITE(6,9)IZ,(DSC(J),J=1,4),A
     9 FORMAT(5H ZONE,I4,2X,4A6,7H, AREA=,F8.3/33X,11HZONE BURDEN)
       CALL HWS(510)
       GO TO 155
C
C      USE PARTS AS ZONES
 180   DO 190 I=1,100                                                    G
       IF(IU(I).EQ.0)GO TO 190
       A=0.
       CALL HES(510,KO)
       DO 185 J=1,4
       IB=IAB(I,J)
       IF(IB.LE.0) GO TO 185
       A=A+DR(IB,1)
       CALL HES(509,IB)
       DR(509,1)=AVF(IB)
       CALL HCS(510,509,510,1)
 185   CONTINUE
       WRITE(6,10)I,DAC(I,1),DAC(I,2),DR(510,1)
   10 FORMAT(5H ZONE,I4,2X,2A6,7H, AREA=,F7.2,8H, BURDEN)                A
       CALL HWS(510)
 190   CONTINUE
       GO TO 100
       END
```

```
      SUBROUTINE MBS(MD,K1,K2)
C     MICROBIAL BUILDUP SUBROUTINE
      COMMON   KK, IK, KR, RUN(7), KS, STG(7), KT, TSK(7), DSC(6), ID,
     .         NE, AES(4), IEC(10), IET(10), AEC(10), AET(10), AED(10),
     .         AEF(10,4), NO, IOT(35), IOQ(35), IOC(35), NP, DAC(120,2),
     .         IU(120), IAB(120,6), AAG(120), AAS(120), AAT(120),
     .         NX(572), DR(572,11), XR(572,11), KO, NS, IR(5), AR(4),
     .         APA(2), NT, JT(100), XMT(100), XVT(100), L1, L2, L3, L4,
     .         KTS, ITE, K3, DUM(2664)
      DIMENSION IX(6),XX(7),DES(6)                                       G
      GO TO (100,120,200), IK                                            G
C
C     ENVIRONMENTS INPUTS-
  100 READ (9) L,IX,AES,(XX(J),J=1,3)
      WRITE(6,8)(AES(J),J=1,4)
    8 FORMAT(//30H ENVIRONMENTS INPUTS - - - - -/5H AES=,4E12.4)
C     AES(J) IS THE SURFACE LIFETIME MODIFIER FOR SURFACE J
      DO 110 I=1,11                                                      G
      READ (9) L,N,(IX(J),J=1,5),DSC,XX(1)
      IF(N.LE.0)GO TO 120
      READ (9) L,IEC(N),IET(N),(IX(J),J=1,4),AEC(N),AET(N),AED(N),
     .(AEF(N,J),J=1,4)
C     IEC,AEC DESCRIBE THE ENVIRONMENTS BIOTA CONCENTRATION
C     IET,AET DESCRIBE THE REFERENCE ACCRETION TIME
C     AED IS THE AIRBOURNE CONCENTRATION PER MAN
C     AEF IS THE RATE AT WHICH, FOR ALL OTHER FACTORS STANDARD,
C        BIOTA REACH EACH SURFACE
      WRITE(6,20)N,(DSC(J),J=1,4),IEC(N),AEC(N),IET(N),AET(N),AED(N),
     .(AEF(N,J),J=1,4)
   20 FORMAT(/I4,1X,4A6,4X,4HIEC=,I3,6H, AEC=,F7.4,6H, IET=,I3,6H, AET=,
     .F7.2,6H, AED=,F7.4,6H, AEF=,4F6.2)
  110 IF(NE.LT.N)NE=N
C
C     OPERATIONS INPUTS-
  120 DO 130 I=1,36                                                      G
      READ (9) L,N,(IX(J),J=1,5),DSC,XX(1)
      IF(N.LE.0)GO TO 200                                                G
      IF ( I .EQ. 1 )          WRITE (6,9)                               G
    9 FORMAT(//30H OPERATIONS INPUTS - - - - - -/)
      READ (9) L,IOT(N),IOQ(N),IOC(N),(IX(J),J=1,3),XX
      WRITE(6,21)N,(DSC(J),J=1,4),IOT(N),IOQ(N),IOC(N)
   21 FORMAT(/I4,1X,4A6,4X,4HIOT=,I3,6H, IOQ=,I3,6H, IOC=,I3)
C     IOT IS THE OPERATION TIME INTERVAL
C     IOQ IS THE DIRTINESS FACTOR
C     IOC IS THE BIOTA CONCENTRATION FOR CONTACT CONTAMINATION
  130 IF(NO.LT.N)NO=N
C
C     SUBTASK LEVEL MICROBIAL BUILDUP DETERMINATION
  200 DO 295 I=1,1001                                                    G
      READ(9) L,NS,L1,L2,L3,IX(1),IX(2),DSC,XX(1)                        G
      IF(NS.LE.0)GO TO 300                                               G
      N1=L1                                                              G
      N2=L2                                                              G
      WRITE(6,12)NS,(DSC(J),J=1,4)                                       G
   12 FORMAT(/13H **** SUBTASK,I3,2H, ,4A6)                              G
C     L1 AND L2 ARE PREREQUISITE TIME DISTRIBUTIONS                      G
C     L3 IS THE FINISH TIME DISTRIBUTION IF NEEDED                       G
C        AS PREREQUISITE FOR ANOTHER SUBTASK                             G
C                                                                        G
C     DETERMINE SUBTASK START TIME                                       G
C     IT IS THE SUBTASK START TIME INDEX                                 G
      IT=0                                                               G
      IF(L1.LE.0)GO TO 210                                               G
```

```
      L1=L1+530                                                         G
      IF(L2.GT.0)GO TO 205                                              G
      IT=L1                                                             G
      GO TO 210                                                         G
  205 L2=L2+530                                                         G
      IT=551                                                            G
      CALL HCS(L1,L2,IT,5)                                              G
      IF(ID.LE.0) GO TO 210                                             G
      IF(NX(IT).GT.NX(ID))CALL HAS(IT)                                  G
  210 IF ( I .NE. 1 )          GO TO 215                                G
C     KTS IS THE TASK START TIME                                        G
      KTS = 552                                                         G
      IF ( IT .EQ. 0 )                    CALL HES(KTS,KO)              G
      IF ( IT .NE. 0 )                    CALL HES(KTS,IT)              G
  215 IF ( IT .NE. 0 )                    GO TO 218                     G
      IT = 551                                                          G
      CALL HES(IT,KTS)                                                  G
  218 LL = IT                                                           G
      T1=DR(IT,1)                                                       G
C
C     PARTS INPUTS-
  140 DO 150 K=1,121                                                    G
      READ (9) L,N,(IX(J),J=1,5),DES,XX(1)                              G
      IF(N.LE.0)GO TO 160
      IF ( K .EQ. 1 .AND. K1 .GE. 1 )   WRITE(6,10)                     G
   10 FORMAT(//30H PARTS INPUTS  - - - - - - - - -/)
      DAC(N,1)=DES(5)                                                   G
      DAC(N,2)=DES(6)                                                   G
      IU(N)=0
      AAT(N)=0.
      IF(IT.GT.0) AAT(N)=DR(IT,1)                                       G
      READ (9) L,(IAB(N,J),J=1,6),AAG(N),AAS(N),(XX(J),J=1,5)
      IF(IAB(N,6).GT.0)GO TO 142
      AAG(N)=0.
      AAS(N)=0.
  142 CONTINUE
C     DAC IS THE PERMANENT ALPHAMERIC DESCRIPTION OF PART N
C     IAB(J) INDICATES THE DISTRIBUTION FOR AREA/BURDEN FOR
C        J=1, TOP SURFACE
C        J=2, OTHER EXTERIOR SURFACE
C        J=3, MATED SURFACE
C        J=4, OCCLUDED SURFACE
C     IAB(5) IS THE ENVIRONMENT INDEX (MAY BE LEFT BLANK)
C     IAB(6) IS THE RETENTION DISTRIBUTION FOR CONTACT
C     AAG IS THE RETENTION DISTRIBUTION MEAN FOR FALLOUT
C     AAS IS THE RETENTION DISTRIBUTION MEAN FOR CONTACT
C     AAT IS THE LAST TIME OF ACCRETION UPDATE
      IF ( K1 .GE. 1 )        WRITE(6,22)  N,(DES(J),J=1,6),(IAB(N,J),  G
     *                        J=1,6),AAG(N),AAS(N)                      G
   22 FORMAT(/I4,1X,6A6,4X,4HIAB=,6I4,6H, AAG=,F7.3,6H, AAS=,F7.3)
      DO 145 J = 1, 4                                                   G
      IF ( IAB(N,J) .GT. 0 )              CALL HES(IAB(N,J),KO)         G
  145 CONTINUE                                                          G
 150  IF(NP.LT.N)NP=N
C
C     DISTRIBUTIONS INPUTS -
  160 DO 175 K=1,501                                                    G
      READ (9) L,N,M,(IX(J),J=1,4),DES,XX(1)                            G
C     M=1 INDICATES THE DISTRIBUTION IS A CONSTANT
      IF(N.LE.0)GO TO 179
      IF ( K .EQ. 1 .AND. K1 .GE. 1 )   WRITE(6,11)                     G
   11 FORMAT(//30H DISTRIBUTIONS INPUTS  - - - -//)
      READ (9) L,(DR(N,J),J=1,11),(XR(N,J),J=1,11)
```

```
      NX(N)=M
      IF(DR(N,1).LE.0.) DR(N,1)=AVF(N)
      IF(N.GT.20) GO TO 171
      J=N+552
      CALL HES(J,N)
  171 IF(ID.LE.0) GO TO 172
      IF(M.GT.NX(ID)) CALL HAS(N)
  172 IF ( K1 .GE. 1 )  WRITE(6,23)  N,(DES(J),J=1,4),DR(N,1)                G
   23 FORMAT(I4,1X,4A6/5X,12HCOEFFICIENT=,E11.4,6X,12HDISTRIBUTION)
      IF ( K1 .GE. 1 )          CALL HWS(N)                                  G
  175 CONTINUE
  179 IF(ID.LE.0) GO TO 190
      IF(MD.EQ.NX(ID)) GO TO 190
      DO 180 K=1,20
      IF(NX(K).GT.NX(ID)) CALL HAS(K)
  180 CONTINUE
      IF(MD.LE.NX(ID)) GO TO 189
      DO 185 K=21,552
      IF(NX(K).GT.NX(ID)) CALL HAS(K)
  185 CONTINUE
  189 MD=NX(ID)
C
  190 M=2                                                                    G
C
      WRITE(11) M,KR,KS,KT,NS,(DSC(J),J=1,4),IT,NX(IT),(DR(IT,J),J=1,        G
     .11),(XR(IT,J),J=1,11),(AEC(J),J=1,5),N1,N2,(JT(J),J=1,3)               G
      NC=0
C     ENVIRONMENT,AREA/BIOTA BURDEN CHANGES-
  220 READ (9) L,K,IR,AR,(XX(J),J=1,3)
C     K IS THE CHANGE IDENTIFIER
C     IR(1),IR(2) ARE THE NEW PART AND SURFACE INDICES
C     IR(3),IR(4) ARE THE CONTRIBUTING PART AND SURFACE
C     IR(5) IS THE NEW ENVIRONMENT (IF ANY) FOR PART IR(1)
C     AR(1) IS THE AREA CHANGED TO IR(1),IR(2) FROM IR(3),IR(4)
C     AR(2) IS THE NEW AAG FACTOR
C     AR(3) IS THE NEW AAS FACTOR
C     IF IR(5), AR(1), AR(2), AR(3) ZERO, NO CHANGE IS MADE
      IF(K.LE.0)GO TO 240
      NC=NC+1
      IF ( NC .EQ. 1 .AND. K1 .GE. 2 )  WRITE(6,16)                          G
   16 FORMAT(/6X,42HENVIRONMENT/AREA/RETENTION FACTOR CHANGES-/)
      I1=IR(1)
      IF(NP.LT.I1)NP=I1
      IF(AR(1).LE.0.)GO TO 232
      CALL MAS(I1,T1)
      IU(I1)=IU(I1)+1
      I2=IR(2)
      I3=IR(3)
      CALL MAS(I3,T1)
      IU(I3)=IU(I3)+1
      I4=IR(4)
      IB=IAB(I1,I2)
      IA=IAB(I3,I4)
      IF(IR(5).GT.0) IAB(I1,5)=IR(5)
      IF(AR(2).GT.0.) AAG(I1)=AR(2)
      IF(AR(3).GT.0.) AAS(I1)=AR(3)
      IF(DR(IA,1).LE.0.)GO TO 220
      F=AR(1)/DR(IA,1)
      IF(F.LT..995)GO TO 230
      F=1.
  230 CALL HMS(IA,F,501)                                                     G
      A=DR(IB,1)+DR(501,1)                                                   G
      DR(IB,1)=AVF(IB)
```

```
      DR(501,1)=AVF(501)                                                G
      CALL HCS(IB,501,IB,1)                                             G
      DR(IB,1)=A
      F=1.-F
      CALL HMS(IA,F,IA)
      GO TO 220
  232 IF(IR(5).GT.0) IAB(I1,5)=IR(5)
      IF(AR(2).GT.0.) AAG(I1)=AR(2)
      IF(AR(3).GT.0.) AAS(I1)=AR(3)
      GO TO 220
C
C     OPERATIONAL LEVEL BUILDUP-
  240 IE=ITE                                                            G
C     IT IS THE CURRENT TIME DISTRIBUTION INDEX                         G
  250 READ (9) L,IO,IKE,(IX(J),J=1,4),AKT,AKQ,(XX(J),J=1,5)
C     IO IS THE OPERATION
C     IKE IS THE OPERATION ENVIRONMENT WHEN DIFFERENT FROM
C        THE TASK ENVIRONMENT OR THE PREVIOUS OPERATION ENVIRONMENT
C     AKT IS THE OPERATION TIME MODIFIER
C     AKQ IS THE NUMBER OF MEN
C        IO LT 0 INDICATES A DECONTAMINATION OPERATION FOR WHICH
C        AKT IS THE FIXED OPERATION INTERVAL
      IF(IO)252,290,258
C
C     DECONTAMINATION
 252  IO=-IO
      T1=DR(IT,1)
      IF(AKT.LE.0.)GO TO 253
      NX(507)=1                                                         G
      DR(507,1)=AKT                                                     G
      XR(507,1)=AKT                                                     G
      CALL HCS(IT,507,IT,1)                                             G
  253 IF(K1.GE.2) WRITE(6,37)IO,T1,DR(IT,1)
   37 FORMAT(5X,9HOPERATION,I3,25H, (DECONTAMINATION), FROM,F8.2,3H TO, G
     .F8.2,6H HOURS)                                                    G
      T1=DR(IT,1)                                                       A
  254 READ (9) L,IP,LK,(IX(J),J=1,4),AR,(XX(J),J=1,3)
C     IP IS THE PART AFFECTED
C     LK IS THE CURVE DESCRIBING FRACTION OF BIOTA REMOVED
C     AR(J) IS THE MEAN FRACTION OF BIOTA REMOVED FROM SURFACE J
      IF(IP.LE.0)GO TO 250
      IU(IP)=IU(IP)+1
      DT=T1-AAT(IP)
      CALL MAS(IP,T1)
      DO 256 J=1,4
      IF(AR(J).LE.0.)GO TO 256
      IB=IAB(IP,J)
      IF(IB.LE.0)GO TO 256
      A=DR(IB,1)
      IF(A.LE.0.) GO TO 256
      IF(DT.LE.0.) GO TO 255
      M=8
C
      WRITE(11)M,KR,KS,KT,NS,(DSC(L),L=1,4),IB,NX(IB),(DR(IB,L),L=1,11),G
     .(XR(IB,L),L=1,11),A,T1,(DAC(IP,L),L=1,2),A,IP,J,IO,IAB(IP,5),IT
  255 F=(1.-AR(J))/DR(LK,1)                                             A
      CALL HMS(LK,F,501)                                                G
      DR(IB,1)=AVF(IB)
      DR(501,1)=AVF(501)                                                G
      IF ( DR(501,1) .LT. 1. )        GO TO 2256                        G
      WRITE (6,2255)       LK
 2255 FORMAT (// 22H *****ERROR***** LK = , I3, 38H INPUT CAUSES NEGATIV   G
     *E DECONTAMINATION //)
```

```
      GO TO 2257
 2256 CALL HCS(IB,501,IB,3)                                              G
 2257 BD=DR(IB,1)
      DR(IB,1)=A
      M=9
C
      WRITE(11)M,KR,KS,KT,NS,(DSC(L),L=1,4),IB,NX(IB),(DR(IB,L),L=1,11),G
     .(XR(IB,L),L=1,11),A,T1,AKT,AR(J),A,IP,J,IO,LK,IP
      IF(K1.EQ.3) WRITE(6,39)IP,DAC(IP,1),DAC(IP,2),J,A,AR(J),BD
   39 FORMAT(10X,4HPART,I4,2X,2A6,9H, SURFACE,I2,7H, AREA=,F8.3,
     .19H, FRACTION REMOVED=,F5.3,17X,9H, BURDEN=,E10.3)                A
 256  CONTINUE
      GO TO 254
C
C     FALLOUT CONTAMINATION-
 258  IF(IKE.GT.0)IE=IKE
      IC=IOC(IO)
      IQ=IOQ(IO)
      II=IET(IE)
      JC=IEC(IE)
      ITK=IOT(IO)
      T1=DR(IT,1)
      IF(ITK.EQ.0)GO TO 259
      F=AKT/DR(ITK,1)
      CALL HMS(ITK,F,507)                                                G
      DR(507,1)=AVF(507)                                                 G
      CALL HCS(IT,507,IT,1)                                              G
C     507 INDICATES THE OPERATION TIME DISTRIBUTION                      G
  259 IF(K1.GE.2) WRITE(6,29) IO,IE,T1,DR(IT,1),AKQ
   29 FORMAT(5X,9HOPERATION,I3,13H, ENVIRONMENT,I3,6H, FROM,F8.2,3H TO, G
     .F8.2,7H HOURS,,F3.0,4H MEN)                                        G
      AQ=AKQ/DR(IQ,1)
      Q = AQ*AED(IE)
      CALL HMS(IQ,Q,505)                                                 G
      DR(505,1)=AVF(505)                                                 G
C     505 IS THE CURVE Q*D                                               G
      AC=AEC(IE)/DR(JC,1)
      CALL HMS(JC,AC,506)                                                G
      DR(506,1)=AVF(506)                                                 G
C     506 IS THE CURVE C                                                 G
      CALL HCS(505,506,506,1)                                            G
C     506 INDICATES THE TOTAL FALLOUT SOURCE CONCENTRATION (C+Q*D)       G
  260 READ (9) L,IP,LS,(IX(J),J=1,4),APD,APC,APS,APA,(XX(J),J=1,2)
C     IP IS THE PART AFFECTED
C     LS IS THE TOOL STICKINESS DISTRIBUTION
C     APD IS NOT USED IN THIS VERSION
C     APC IS THE TOOL BIOTA CONCENTRATION MODIFIER
C     APS IS THE MEAN TOOL STICKINESS
C     APA IS THE CONTACT AREA FOR EACH TOUCHED SURFACE
      IF(IP.LE.0)GO TO 250
      IU(IP)=IU(IP)+1
      DT=T1-AAT(IP)
      CALL MAS(IP,T1)
      IF(AQ.LE.0.)GO TO 275
      IF(IE.LE.0)GO TO 275
      IF(ITK.LE.0)GO TO 275
      AAT(IP)=DR(IT,1)
      DO 270 J=1,4
      IB=IAB(IP,J)
      F=AEF(IE,J)*AAG(IP)
      IF(IB.LE.0)GO TO 270
      IF(F.LE.0.)GO TO 270
      A=DR(IB,1)
```

```
      IF(A.LE.0.)GO TO 270
      IF(DT.LE.0.) GO TO 263
      M=8
C
      WRITE(11)M,KR,KS,KT,NS,(DSC(L),L=1,4),IB,NX(IB),(DR(IB,L),L=1,11),G
     .(XR(IB,L),L=1,11),A,T1,(DAC(IP,L),L=1,2),A,IP,J,IO,IAB(IP,5),IT
  263 F=F*A
      CALL HMS(506,F,505)                                               G
C     505 IS THE CURVE A*R = A*F*G*(C+Q*D)                              G
      EE=AET(IE)*AES(J)/DR(II,1)
      NNNC = NX(II)
      NNNCP1 = NNNC + 1
      DO 2264 IJK = 1, NNNC
      XR(502,IJK) = 1. / ( XR(II,NNNCP1-IJK) * EE )
      IF ( IJK .EQ. NNNC )         GO TO 2264
      DR(502,IJK+1) = DR(II,NNNCP1-IJK)
 2264 CONTINUE
      NX(502) = NNNC
      DR(502,1)=AVF(502)                                                G
C     502 IS THE CURVE 1 / V
      CALL HCS(505,II,504,3)
      CALL HMS(504,EE,504)
C     504 IS THE CURVE A*V*R                                            G
      CALL HCS(507,502,502,3)
C     502 IS THE CURVE T/V                                              G
      M=NX(502)                                                         G
      NX(503)=M                                                         G
      MP1 = M + 1
      DO 265 JJ=1,M
      XR(503,JJ)=EXP(-XR(502,MP1-JJ))
      IF ( JJ .EQ. M )       GO TO 265
      DR(503,JJ+1) = DR(502,MP1-JJ)
  265 CONTINUE
C     503 IS THE CURVE EXP(-T/V)
      DO 266 JJ = 1, M
      XR(502,JJ) = 1. - XR(503,MP1-JJ)
      IF ( JJ .EQ. M )          GO TO 266
      DR(502,JJ+1) = DR(503,MP1-JJ)
  266 CONTINUE
C     502 IS THE CURVE 1-EXP(-T/V)
      DR(502,1)=AVF(502)                                                G
      DR(503,1)=AVF(503)                                                G
      CALL HCS(504,502,504,3)
      DR(IB,1)=AVF(IB)
      CALL HCS(IB,503,503,3)
      CALL HCS(503,504,IB,1)
      CALL HAS(IB)
      BD=DR(IB,1)
      DR(IB,1)=A
      M=3
C
      WRITE(11)M,KR,KS,KT,NS,(DSC(L),L=1,4),IB,NX(IB),(DR(IB,L),L=1,11),G
     .(XR(IB,L),L=1,11),A,T1,AKT,AKQ,A,IP,J,IO,IE,IP
      IF(K1.EQ.3) WRITE(6,30) IP,DAC(IP,1),DAC(IP,2),J,A,APC,APA(J),BD
   30 FORMAT(10X,4HPART,I4,2X,2A6,9H, SURFACE,I2,7H, AREA=,F8.3,
     .14H, TOOL BURDEN=,F6.0,15H, AREA TOUCHED=,F6.3,9H, BURDEN=,E10.3)
  270 CONTINUE
C
C     CONTACT CONTAMINATION-
  275 IF(IC.LE.0)GO TO 260
      DR(IC,1)=AVF(IC)
      IG=IAB(IP,6)
      IF(IG.LE.0) GO TO 260                                             A
```

```
      DR(IG,1)=AVF(IG)
      S1=.5*AAS(IP)*APC
      IF(LS.LE.0)GO TO 260
      DR(LS,1)=AVF(LS)
      IF(DR(LS,1).LE.0.)GO TO 260
      S2=.5*APS/DR(LS,1)
      B=S1*S2
      IF(B.EQ.0.)GO TO 260
      S1=S1/(DR(IC,1)*DR(IG,1))
      DO 280 J=1,2
      IB=IAB(IP,J)
      IF(IB.LE.0)GO TO 280
      A=DR(IB,1)
      IF(A.LE.0.)GO TO 280
      IF(APA(J).LE.0.)GO TO 280
      DR(IB,1)=AVF(IB)
      F=S1*APA(J)
      CALL HMS(IC,F,504)                                                G
      CALL HCS(504,IG,504,3)                                            G
      F=S2*APA(J)/A
      NNNC = NX(LS)
      NNNCP1 = NNNC + 1
      DO 2280 IJK = 1, NNNC
      XR(505,IJK) = 1. - XR(LS,NNNCP1-IJK) * F
      IF ( IJK .EQ. NNNC )          GO TO 2280
      DR(505,IJK+1) = DR(LS,NNNCP1-IJK)
 2280 CONTINUE
      NX(505) = NNNC
      DR(505,1) = AVF(505)
      CALL HCS(IB,505,IB,3)
      CALL HCS(IB,504,IB,1)                                             G
      CALL HAS(IB)
      DR(IB,1)=A
      M=4
C
      WRITE(11)M,KR,KS,KT,NS,(DSC(L),L=1,4),IB,NX(IB),(DR(IB,L),L=1,11),G
     .(XR(IB,L),L=1,11),A,APA(J),AAS(IP),APS,APC,IP,J,IAB(IP,6),IC,IP
 280  CONTINUE
      GO TO 260
 290  IF(DR(LL,1).LT.DR(IT,1))LL=IT
      IF(L3.LE.0)GO TO 295                                              G
      L3=L3+530                                                         G
      CALL HES(L3,LL)                                                   G
 295  CONTINUE
  300 T=DR(LL,1)                                                        G
      WRITE(6,601) KR,KS,KT,(TSK(J),J=1,7)
  601 FORMAT(13H1TASK SUMMARY/4H RUN,I2,7H, STAGE,I3,6H, TASK,I4,2X,7A6,
     .//29X,15HTASK START TIME)
      CALL HWS(KTS)
      WRITE(6,604)
  604 FORMAT(29X,16HTASK FINISH TIME)
      CALL HWS(LL)
      IF(K3.NE.0) WRITE(6,605)                                          G
  605 FORMAT(/36H BURDEN DISTRIBUTIONS AT END OF TASK//50H BURDENS BY ZO
     .NE AND SURFACE    - - - - - - - - - - -/)
      DO 303 J=1,4
      JJ=503+J                                                          G
      CALL HES(JJ,KO)
  303 AR(J)=0.
      DO 310 I=1,100                                                    G
      IF(IU(I).EQ.0)GO TO 310
      CALL MAS(I,T)
      IF(K3.NE.0) WRITE(6,602) I,(DAC(I,J),J=1,2)                       G
```

```
 602 FORMAT(/5H ZONE,I4,2X,2A6)
     DO 305 J=1,4
     JJ=503+J                                                           G
     IB=IAB(I,J)
     IF(IB.LE.0)GO TO 305
     IF(DR(IB,1).LE.0.)GO TO 305
     A=DR(IB,1)
     AR(J)=AR(J)+A
     DR(IB,1)=AVF(IB)
     CALL HCS(JJ,IB,JJ,1)
     DR(IB,1)=A
     IF(K3.NE.0) WRITE(6,603) J,DR(IB,1)                                G
 603 FORMAT(5X,14HSURFACE NUMBER,I2,7H, AREA=,F9.3,8H, BURDEN)
     IF ( K3 .NE. 0 )         CALL HWS(IB)                              G
305  CONTINUE
310  CONTINUE
     WRITE(6,608)
 608 FORMAT(//50H BURDEN TOTALS BY SURFACE - - - - - - - - - - - - - -//)
     A=0.
     CALL HES(501,K0)                                                   G
     DO 320 J=1,4
     JJ=503+J                                                           G
     CALL HAS(JJ)
     WRITE(6,606) J,AR(J)
 606 FORMAT(/8H SURFACE,I2,13H, TOTAL AREA=,F9.3,14H, TOTAL BURDEN)
     CALL HWS(JJ)
     A=A+AR(J)
320  CALL HCS(501,JJ,501,1)                                             G
     CALL HAS(501)                                                      G
     WRITE(6,609)
 609 FORMAT(//50H - - - - - - - - - - - - - - - - - - - - - - - - -//)
     WRITE(6,607) A
 607 FORMAT(//28H TOTAL, ALL SURFACES.  AREA=,F9.3,9H,  BURDEN)
     CALL HWS(501)                                                      G
     JT(NT)=KT
     XMT(NT)=AVF(501)                                                   G
     XVT(NT)=AVF(LL)
     IF(L4.LE.0)RETURN
     L4=L4+530
     CALL HES(L4,501)                                                   G
     DR(L4,1)=A
     RETURN
     END
```

```
      SUBROUTINE HCS(IA,IB,IC,K)
C
C     HISTOGRAM COMBINING SUBROUTINE
      DIMENSION CA(11),CB(11),CR(21),ZR(21),QR(11,11),NQ(11),XCORR(3)
      COMMON   KK, IK, KR, RUN(7), KS, STG(7), KT, TSK(7), DSC(6), ID,
     .         NE, AES(4), IEC(10), IET(10), AEC(10), AET(10), AED(10),
     .         AEF(10,4), NO, IOT(35), IOQ(35), IOC(35), NP, DAC(120,2),
     .         IU(120), IAB(120,6), AAG(120), AAS(120), AAT(120),
     .         NX(572), DR(572,11), XR(572,11), KO, NS, IR(5), AR(4),
     .         APA(2), NT, JT(100), XMT(100), XVT(100), L1, L2, L3, L4,
     .         KTS, ITE, K3, DUM(2664)
      IF ( K .NE. 3 )             GO TO 20
      XR(IA,1) = AMAX1 ( XR(IA,1), 1.E-4 )
      XR(IB,1) = AMAX1 ( XR(IB,1), 1.E-4 )
   20 CONTINUE
      NA=NX(IA)
      NB=NX(IB)
      A=DR(IA,1)
      B=DR(IB,1)
      X=XR(IA,1)
      Y=XR(IB,1)
      CR(1)=ZF(A,B,K)
      CM=CR(1)
      DR(IC,1)=CR(1)
      ZR(1)=ZF(X,Y,K)
      IF(NA.GT.1)GO TO 50
      IF(X.NE.0.)GO TO 30
      IF((K-3)*(K-4).EQ.0)GO TO 35
   30 IF(NB.GT.1)GO TO 40
   35 XR(IC,1)=ZR(1)
      NX(IC)=1
      GO TO 1000
   40 NC=NB
      DO 45 J=2,NC
      Y=XR(IB,J)
      CR(J)=DR(IB,J)
   45 ZR(J)=ZF(X,Y,K)
      GO TO 60
   50 IF(NB.GT.1)GO TO 90
      NC=NA
      DO 55 J=2,NC
      X=XR(IA,J)
      CR(J)=DR(IA,J)
   55 ZR(J)=ZF(X,Y,K)
   60 NX(IC)=NC
      IF(ZR(NC).GT.ZR(1))GO TO 70
      XR(IC,1)=ZR(NC)
      JC=NC
      DO 65 J=2,NC
      DR(IC,J)=CR(JC)
      JC=JC-1
   65 XR(IC,J)=ZR(JC)
      GO TO 80
   70 XR(IC,1)=ZR(1)
      DO 75 J=2,NC
      DR(IC,J)=CR(J)
   75 XR(IC,J)=ZR(J)
   80 IF(K.LE.4)GO TO 1000
C     ELIMINATION OF DUPLICATE X VALUES FOR K=5
      CR(1)=0.
      CR(2)=0.
      JC=1
      ZR(1)=XR(IC,1)
```

```
      DO 85 J=2,NC
      CR(J)=CR(J)+DR(IC,J)
      IF(XR(IC,J).LE.XR(IC,J-1))GO TO 85
      JC=JC+1
      CR(JC)=CR(J)
      ZR(JC)=XR(IC,J)
 85   CR(J+1)=CR(J)
      NC=JC
      GO TO 300
 90   IF(K.EQ.5)GO TO 200
      JA=0
      JB=0
      JC=0
      KA=NA+1
      KB=NB+1
      IF(NA.GE.NB)GO TO 94
      NC=NB
      KC=KB
      GO TO 100
 94   NC=NA
      KC=KA
100   JC=JC+1
      KC=KC-1
      IF(JC.GT.KC)GO TO 140
      IF(JA+1.GE.KA)GO TO 124
      JA=JA+1
      KA=KA-1
      X1=XR(IA,JA)
      X2=XR(IA,KA)
      IF(JB+1.GE.KB)GO TO 120
      JB=JB+1
      KB=KB-1
      Y1=XR(IB,JB)
      Y2=XR(IB,KB)
      IF(K.EQ.2)GO TO 110
      Z1=ZF(X1,Y1,K)
      Z2=ZF(X1,Y2,K)
      IF(Z2.GT.Z1)GO TO 102
      ZR(JC)=Z2
      GO TO 104
102   ZR(JC)=Z1
104   Z1=ZF(X2,Y1,K)
      Z2=ZF(X2,Y2,K)
      IF(Z2.GT.Z1)GO TO 106
      ZR(KC)=Z1
      GO TO 100
106   ZR(KC)=Z2
      GO TO 100
110   ZR(JC)=ZF(X1,Y1,K)
      ZR(KC)=ZF(X2,Y2,K)
      IF(ZR(JC)-ZR(KC))100,112,112
112   ZR(JC)=.5*(ZR(JC)+ZR(KC))
114   KC=KC+1
      IF(KC.GT.NC)GO TO 116
      JC=JC+1
      ZR(JC)=ZR(KC)
      GO TO 114
116   NC=JC
      GO TO 140
120   Y1=.5*(XR(IB,JB)+XR(IB,KB))
      ZR(JC)=ZF(X1,Y1,K)
      ZR(KC)=ZF(X2,Y1,K)
      GO TO 100
```

```
  124   X1=.5*(XR(IA,JA)+XR(IA,KA))
        IF(JB+1.GE.KB)GO TO 128
        JB=JB+1
        KB=KB-1
        Y1=XR(IB,JB)
        Y2=XR(IB,KB)
        IF(K.EQ.2)GO TO 130
        Z1=ZF(X1,Y1,K)
        Z2=ZF(X1,Y2,K)
        IF(Z2.GT.Z1)GO TO 126
        ZR(JC)=Z2
        ZR(KC)=Z1
        GO TO 100
  126   ZR(JC)=Z1
        ZR(KC)=Z2
        GO TO 100
  128   Y1=.5*(XR(IB,JB)+XR(IB,KB))
        ZR(JC)=ZF(X1,Y1,K)
        ZR(KC)=ZR(JC)
        GO TO 100
  130   ZR(JC)=ZF(X1,Y2,K)
        ZR(KC)=ZF(X1,Y1,K)
        GO TO 100
  140   NX(IC)=NC
        IF(NC.EQ.1)GO TO 35
        DO 150 J=1,NC
  150   CR(J)=0.
        DO 170 JA=2,NA
        DO 170 JB=2,NB
        P=DR(IA,JA)*DR(IB,JB)
        X1=XR(IA,JA-1)
        X2=XR(IA,JA)
        Y1=XR(IB,JB-1)
        Y2=XR(IB,JB)
        Z1=ZF(X1,Y1,K)
        Z2=ZF(X1,Y2,K)
        IF(Z2.GT.Z1)GO TO 154
        ZA=Z2
        GO TO 155
  154   ZA=Z1
  155   Z1=ZF(X2,Y1,K)
        Z2=ZF(X2,Y2,K)
        IF(Z2.GT.Z1)GO TO 158
        ZB=Z1
        GO TO 160
  158   ZB=Z2
  160   DO 170 JC=2,NC
        IF(ZR(JC).LE.ZA)GO TO 170
        IF(ZR(JC).GE.ZB)GO TO 166
        YC = XF ( ZR(JC), X1, K )
        XC = XF ( ZR(JC), Y1, K )
        IF ((X1.EQ.0..OR.Y1.EQ.0.) .AND. K.EQ.3 )      GO TO 2000
        X4 = AMIN1 ( X2, XC )
        Y4 = AMIN1 ( Y2, YC )
        XTOT = ( X2 - X1 ) * ( Y2 - Y1 )
        X3 = X1
        Y3 = Y1
C
        IF ( YC .GT. Y2 )                        X3 = XF ( ZR(JC), Y2, K )
        IF ( XC .GT. X2 )                        Y3 = XF ( ZR(JC), X2, K )
        XSQUAR = ( Y3 - Y 1 ) * ( X3 - X1 )
C
        TRIANG = .5 * (X4-X3) * (Y4-Y3)
```

```
      XFRAC = (Y2-Y1) * (X3-X1) + (X2-X1) * (Y3-Y1) + TRIANG - XSQUAR
      CR(JC) = CR(JC) + P * XFRAC / XTOT
      GO TO 170
C
 166  CR(JC)=CR(JC)+P
  170 CONTINUE
      GO TO 300
C
 200  JA=1
      JB=1
      JC=0
      KA=0
      KB=0
      CA(1)=0.
      CB(1)=0.
      NM=NA
      IF(NM.LT.NB)NM=NB
 205  IF(XR(IA,JA)-XR(IB,JB))210,235,255
 210  IF(KA.GT.0)GO TO 260
      IF(XR(IA,JA).LT.XR(IB,1))GO TO 225
      DB=CB(JB-1)+(CB(JB)-CB(JB-1))*(XR(IA,JA)-XR(IB,JB-1))/
     .(XR(IB,JB)-XR(IB,JB-1))
      GO TO 220
 215  DB=CB(NB)
 220  JC=JC+1
      CR(JC)=CA(JA)*DB
      ZR(JC)=XR(IA,JA)
 225  IF(JA.GE.NA)GO TO 230
      JA=JA+1
      CA(JA)=CA(JA-1)+DR(IA,JA)
      GO TO 205
 230  IF(KB.GT.0)GO TO 280
      KA=1
      GO TO 205
 235  JC=JC+1
      CR(JC)=CA(JA)*CB(JB)
      ZR(JC)=XR(IA,JA)
      IF(JA.GE.NA)GO TO 245
      JA=JA+1
      CA(JA)=CA(JA-1)+DR(IA,JA)
 240  IF(JB.GE.NB)GO TO 250
      JB=JB+1
      CB(JB)=CB(JB-1)+DR(IB,JB)
      GO TO 205
 245  IF(KB.GT.0)GO TO 280
      KA=1
      GO TO 240
 250  IF(KA.GT.0)GO TO 280
      KB=1
      GO TO 205
 255  IF(KB.GT.0)GO TO 215
      IF(XR(IB,JB).LT.XR(IA,1))GO TO 270
      DA=CA(JA-1)+(CA(JA)-CA(JA-1))*(XR(IB,JB)-XR(IA,JA-1))/
     .(XR(IA,JA)-XR(IA,JA-1))
      GO TO 265
 260  DA=CA(NA)
 265  JC=JC+1
      CR(JC)=CB(JB)*DA
      ZR(JC)=XR(IB,JB)
 270  IF(JB.GE.NB)GO TO 275
      JB=JB+1
      CB(JB)=CB(JB-1)+DR(IB,JB)
      GO TO 205
```

```
 275  IF(KA.GT.0)GO TO 280
      KB=1
      GO TO 205
 280  NC=JC
 290  IF(NC.LE.NM)GO TO 300
      DJ=1.
      JJ=2
      NC=NC-1
      DO 294 J=2,NC
      DS=(CR(J-1)+(CR(J+1)-CR(J-1))*(ZR(J)-ZR(J-1))/(ZR(J+1)-ZR(J-1))
     .-CR(J))**2
      IF(DS.GE.DJ)GO TO 294
      DJ=DS
      JJ=J
      IF(DJ.LT..0001)GO TO 296
 294  CONTINUE
 296  DO 298 J=JJ,NC
      CR(J)=CR(J+1)
 298  ZR(J)=ZR(J+1)
      GO TO 290
  300 XR(IC,1)=ZR(1)
 308  DO 310 J=2,NC
      DR(IC,J)=CR(J)-CR(J-1)
 310  XR(IC,J)=ZR(J)
      NX(IC)=NC
C
      IF ( NC .LT. 4 )                       GO TO 900
      XR(IC,1) = XR(IC,2) - 0.66667 * (XR(IC,2)-XR(IC,1))
      XR(IC,NC) = XR(IC,NC-1) + 0.66667 * (XR(IC,NC)-XR(IC,NC-1))
C
      XCORR(2) = XR(IC,1) / XR(IC,2)
      XCORR(3) = XR(IC,NC) / XR(IC,NC-1)
C
  325 JD = NC - 1
      SUM1 = 0.
      DO 400 J = 3, JD
      SUM1 = SUM1 + DR(IC,J) * ( XR(IC,J-1) + XR(IC,J) )
  400 CONTINUE
      SUM = .5 * (DR(IC,2)*XR(IC,1) + DR(IC,NC)*XR(IC,NC))
      SUM1 = .5*(SUM1+DR(IC,2)*XR(IC,2)+DR(IC,NC)*XR(IC,JD))
      XCORR(1) = ( CM - SUM ) / SUM1
      IFLAG = 1
      IF ( XCORR(1) .LT. XCORR(2) )       IFLAG = 2
      IF ( XCORR(1) .GT. XCORR(3) )       IFLAG = 3
      DO 800 J = 2, JD
      XR(IC,J) = XR(IC,J) * XCORR(IFLAG)
  800 CONTINUE
C
      IF ( (XR(IC,1)+0.001) .LT. XR(IC,2) )       GO TO 850
      DO 825 J = 2, NC
      DR(IC,J-1) = DR(IC,J)
      XR(IC,J-1) = XR(IC,J)
  825 CONTINUE
      DR(IC,2) = DR(IC,2) + DR(IC,1)
      NC = NC - 1
      NX(IC) = NC
      IFLAG = 2
C
  850 IF ( (XR(IC,NC-1)+.001) .LT. XR(IC,NC) )       GO TO 875
      DR(IC,NC-1) = DR(IC,NC-1) + DR(IC,NC)
      NX(IC) = NC - 1
      IFLAG = 2
C
```

```
  875 IF ( IFLAG .EQ. 1 )                GO TO 1000
  900 F = AVF(IC)
      IF ( F .EQ. 0. )          GO TO 1000
      F = CM / F
      CALL HMS (IC,F,IC)
C
 1000 DR(IC,1)=AVF(IC)
      RETURN
 2000 CALL HWS(IA)
      CALL HWS(IB)
      CALL ABNORML
      END
```

```
      FUNCTION XF ( Z, XY, K )
C
      GO TO ( 100, 200, 300, 400 ), K
C
  100 XF = Z - XY
      RETURN
  200 XF = Z + XY
      RETURN
  300 XF = Z / XY
      RETURN
  400 XF = Z * XY
      RETURN
      END
```

Appendix B

COMPUTER PROGRAM LISTING OF CONTROL CARDS FOR DETAILED CHECKING

```
      PROGRAM CHECK ( INPUT, OUTPUT, TAPE5 = INPUT, TAPE6 = OUTPUT )
C
C      CONTROL PROGRAM TO EXERCISE HISTOGRAM MANIPULATION ROUTINES
C
      COMMON   KK, IK, KR, RUN(7), KS, STG(7), KT, TSK(7), DSC(6), ID,
     .         NE, AES(4), IEC(10), IET(10), AEC(10), AET(10), AED(10),
     .         AEF(10,4), NO, IOT(35), IOQ(35), IOC(35), NP, DAC(120,2),
     .         IU(120), IAB(120,6), AAG(120), AAS(120), AAT(120),
     .         NX(572), DR(572,11), XR(572,11), KO, NS, IR(5), AR(4),
     .         APA(2), NT, JT(100), XMT(100), XVT(100), L1, L2, L3, L4,
     .         KTS, ITE, K3, DUM(2664)
      L = 1
C
  100 READ (5,125)       INXX,   IKON,   NTIMES,   KTS
  125 FORMAT ( 4I10 )
      IF ( EOF, 5 ) 150, 175
  150 STOP
C
  175 NX(1) = INXX
      READ (5,200)       ( DR(1,J), J = 1, INXX )
      READ (5,200)       ( XR(1,J), J = 1, INXX )
  200 FROMAT ( 8E10.3 )
C
      DR(1,1) = AVF(1)
      CALL HWS (1)
      IF ( KTS .EQ. 0 )                      GO TO 210
      READ (5,125)       INXX
      NX(2) = INXX
      READ (5,200)       ( DR(2,J), J = 1, INXX )
      READ (5,200)       ( XR(2,J), J = 1, INXX )
      DR(2,1) = AVF(2)
      CALL HWS (2)
       GO TO 220
C
  210 CALL HES (2,1)
  220 NTIMES = NTIMES - 1
C
      DO 225 I = 1, NTIMES
      CALL HCS (1,2,2,IKON)
      CALL HWS (2)
  225 CONTINUE
       GO TO 100
C
      END
```

Appendix C

SAMPLE CALCULATIONS

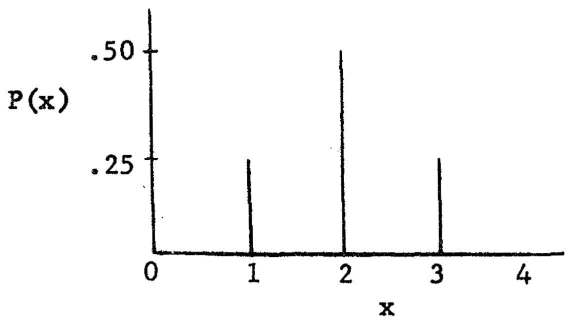
In this appendix are samples of the types of calculations discussed in the main body of the report, and quantitative support for some of the logic that was presented in general terms. In order to keep the calculations within reasonable limits, the histograms used are two- and three-bump whereas ten-bump is most common in the MBPM, with an occasional use of fewer bumps. It is hoped that these calculations will aid in understanding the points discussed in the report.

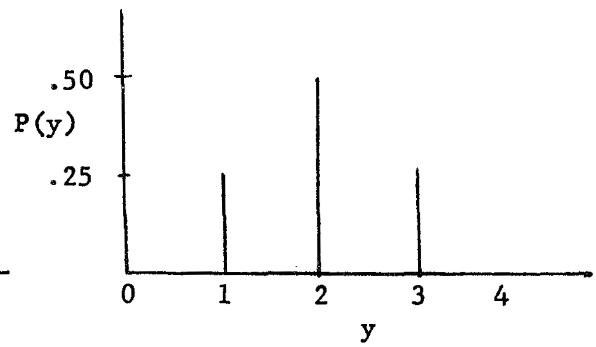
1. Sum of two equal discrete variables

Given that: $P(x=1) = P(y=1) = 0.25$

$P(x=2) = P(y=2) = 0.50$

$P(x=3) = P(y=3) = 0.25$



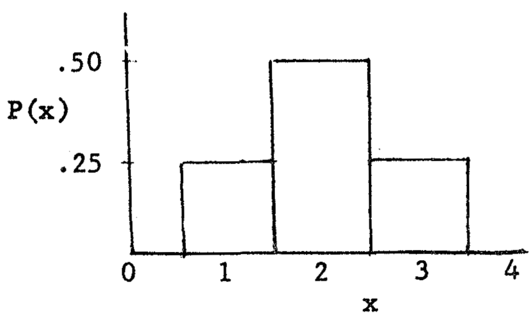


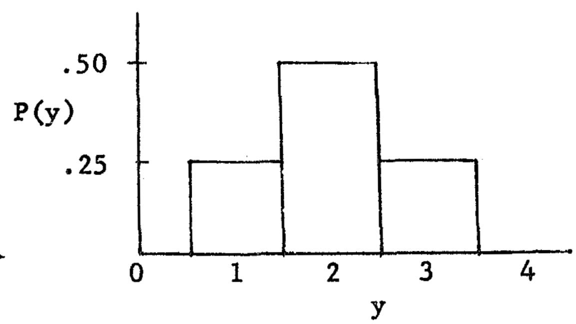
Expressed as histograms:

$P(0.5 \leq x < 1.5) = P(0.5 \leq y < 1.5) = 0.25$

$P(1.5 \leq x < 2.5) = P(1.5 \leq y < 2.5) = 0.50$

$P(2.5 \leq x < 3.5) = P(2.5 \leq y < 3.5) = 0.25$





If $z = x + y$, the joint probabilities are:

$P(0.5 \leq x < 1.5, \ 0.5 \leq y < 1.5) = (0.25)(0.25) = 0.0625$

$P(1.5 \leq x < 2.5, \ 0.5 \leq y < 1.5) = (0.50)(0.25) = 0.125$

$P(2.5 \leq x < 3.5, \ 0.5 \leq y < 1.5) = (0.25)(0.25) = 0.0625$

$P(0.5 \leq x < 1.5,\ 1.5 \leq y < 2.5) = (0.25)(0.50) = 0.125$

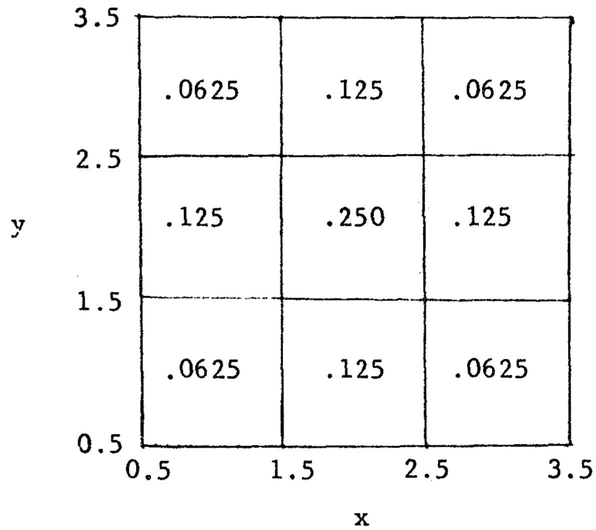
$P(1.5 \leq x < 2.5,\ 1.5 \leq y < 2.5) = (0.50)(0.50) = 0.250$

$P(2.5 \leq x < 3.5,\ 1.5 \leq y < 2.5) = (0.25)(0.50) = 0.125$

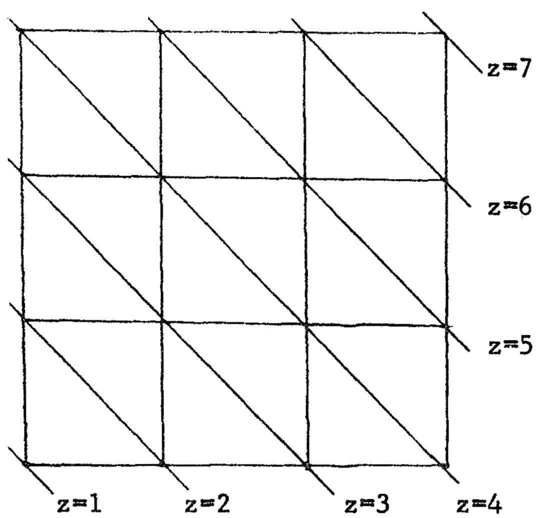
$P(0.5 \leq x < 1.5,\ 2.5 \leq y < 3.5) = (0.25)(0.25) = 0.0625$

$P(1.5 \leq x < 2.5,\ 2.5 \leq y < 3.5) = (0.50)(0.25) = 0.125$

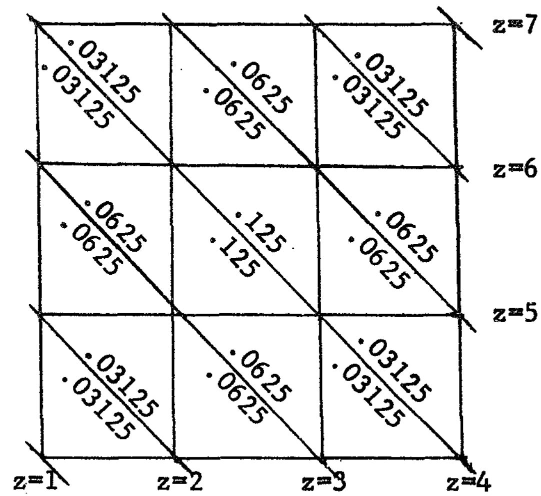
$P(2.5 \leq x < 3.5,\ 2.5 \leq y < 3.5) = (0.25)(0.25) = 0.0625$



The lines of constant z are



Assuming the probability to be proportional to the area gives:



And the distribution of z is

$P(1 \leq x < 2) = 0.03125$

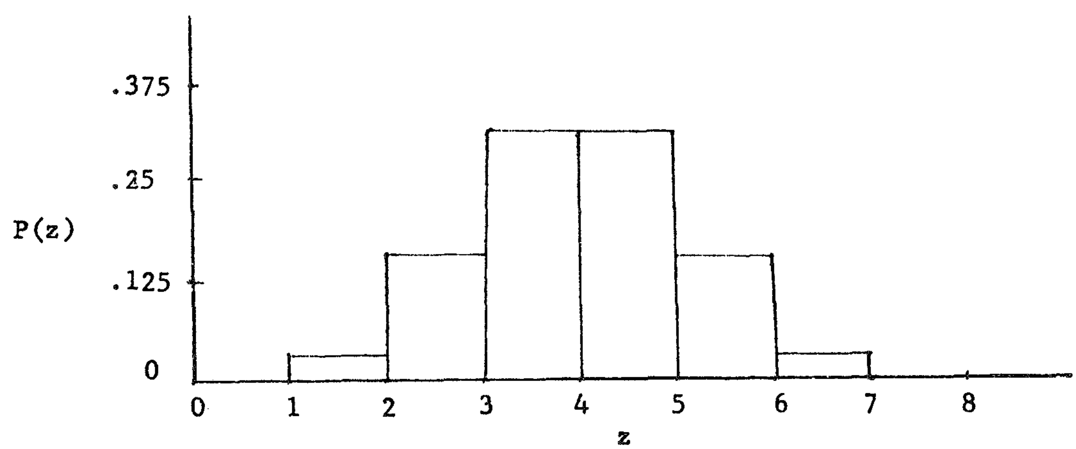
$P(2 \leq z < 3) = .0625 + .03125 + .0625 = 0.15625$

$P(3 \leq z < 4) = .03125 + .0625 + .125 + .0625 + .03125 = 0.3125$

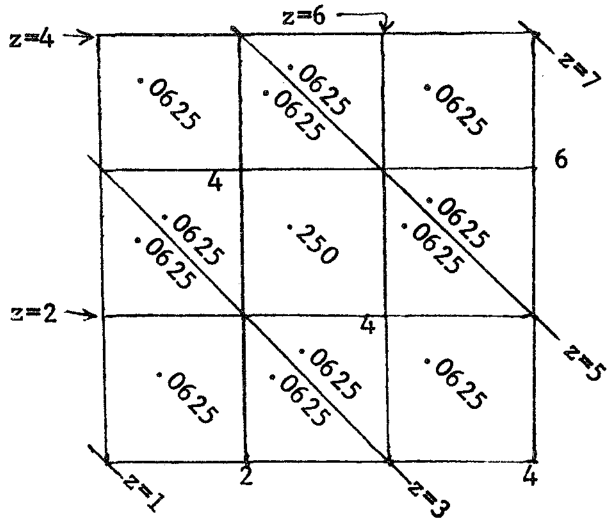
$P(4 \leq z < 5) = .03125 + .0625 + .125 + .0625 = .03125 = 0.3125$

$P(5 \leq z < 6) = .0625 + .03125 + .0625 = 0.15625$

$P(6 \leq z < 7) = 0.03125$



The computer program would calculate this as



$$P(1 \leq z < 3) = .0625 + \frac{3-2}{4-2}(.125) + \frac{3-2}{4-2}(.125) = 0.1875$$

$$P(3 \leq z < 5) = .0625 + .25 + .0625 + \frac{4-3}{4-2}(.125)(2) + \frac{5-4}{6-4}(.125)(2)$$

$$= 0.625$$

$$P(5 \leq z < 7) = .0625 + \frac{6-5}{6-4}(.125)(2) = 0.1875$$

$$\bar{z} = 4.0$$

While the sum of the original discrete distributions is

$P(x=1,\ y=1) = 0.0625$

$P(x=2,\ y=1) = 0.125$

$P(x=3,\ y=1) = 0.0625$

$P(x=1,\ y=2) = 0.125$

$P(x=2,\ y=2) = 0.25$

$P(x=3,\ y=2) = 0.125$

$P(x=1,\ y=3) = 0.0625$

$P(x=2,\ y=3) = 0.125$

$P(x=3,\ y=3) = 0.0625$

$P(z=2) = 0.0625$

$P(z=3) = 0.25$

$P(z=4) = 0.375$

$P(z=5) = 0.25$

$P(z=6) = 0.0625$

The means of the original distributions are

$$\bar{x} = \bar{y} = (.25)(1) + (.50)(2) + (.25)(3) = 2.0$$

So that $(\bar{x})(\bar{y}) = 4.0$

The mean of the sum is

$$\bar{z} = (.0625)(2) + (.25)(3) + (.375)\ (4) + (.25)(5) + (.0625)(6)$$

$$= 4.0$$

And the mean of the histogram of z is

$$\bar{z} = (.03125)\left(\frac{1+2}{2}\right) + (.15625)\left(\frac{2+3}{2}\right) + (.3125)\left(\frac{3+4}{2}\right) + (.3125)\left(\frac{4+5}{2}\right)$$

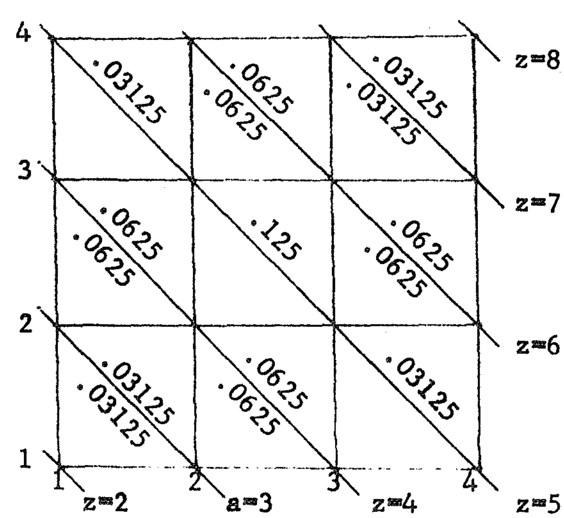
$$+ (.15625)\left(\frac{5+6}{2}\right) + (.03125)\left(\frac{6+7}{2}\right) = 4.0$$

When the histogram intervals are selected at other points for the same distribution

$P(1.0 \leq x < 2.0) = P(1.0 \leq y < 2.0) = 0.25$

$P(2.0 \leq x < 3.0) = P(2.0 \leq y < 3.0) = 0.50$

$P(3.0 \leq x < 4.0) = P(3.0 \leq y < 4.0) = 0.25$



$P(2 \le z < 3) = 0.03125$

$P(3 \le z < 4) = 0.15625$

$P(4 \le z < 5) = 0.3125$

$P(5 \le z < 6) = 0.3125$

$P(6 \le z < 7) = 0.15625$

$P(7 \le z < 8) = 0.03125$

$\bar{z} = 5.0$

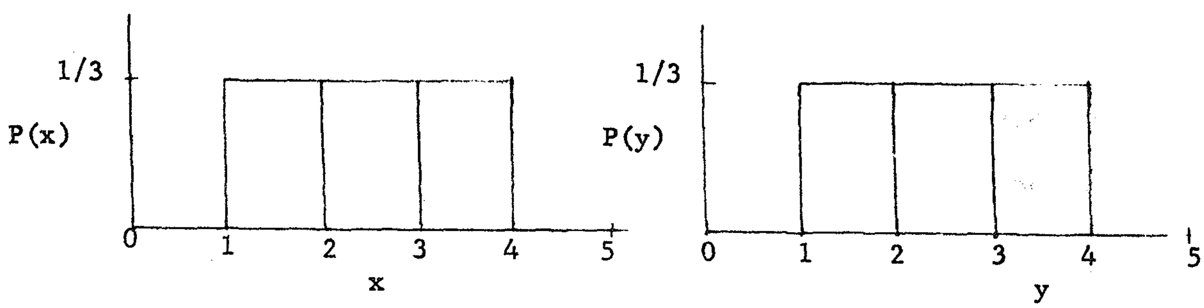
2. <u>Sum of two equal uniformly distributed variables</u>

Given that:

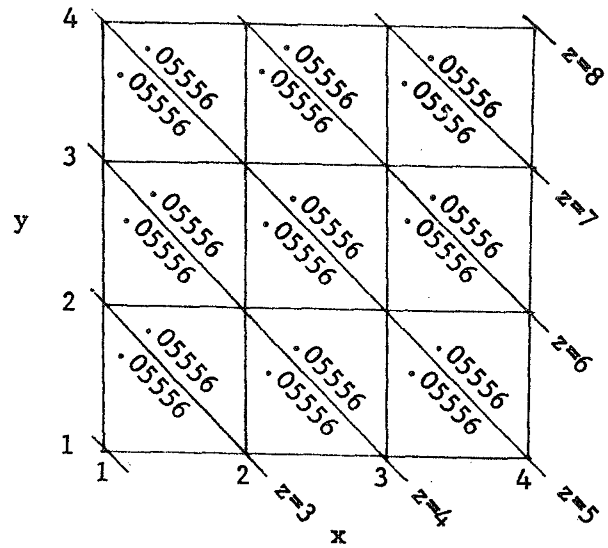
$P(1 \le x < 2) = P(1 \le y < 2) = 1/3$

$P(2 \le x < 3) = P(2 \le y < 3) = 1/3$

$P(3 \le x < 4) = P(3 \le y < 4) = 1/3$



For $z = x + y$



$P(2 \leq z < 3) = 0.05556$

$P(3 \leq z < 4) = 0.16667$

$P(4 \leq z < 5) = 0.27778$

$P(5 \leq z < 6) = 0.27778$

$P(6 \leq z < 7) = 0.16667$

$P(7 \leq z < 8) = 0.05556$

$\bar{x} + \bar{y} = 2.5 + 2.5 = 5.0$

$\bar{z} = 5.0$

If the same distribution were represented by a single bump

$P(1 \leq x < 4) = P(1 \leq y < 4) = 1.0$

which gives a single bump for z

$P(2 \leq z < 8) = 1.0$

inferring that

$P(2 \leq z < 3) = 1/6$

$P(3 \leq z < 4) = 1/6$

$P(4 \leq z < 5) = 1/6$

$P(5 \leq z < 6) = 1/6$

$P(6 \leq z < 7) = 1/6$

$P(7 \leq z < 8) = 1/6$

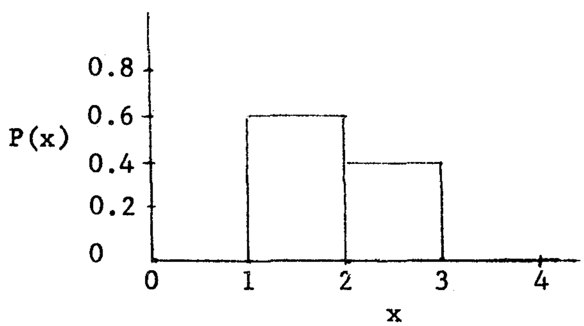
which is quite a different distribution.

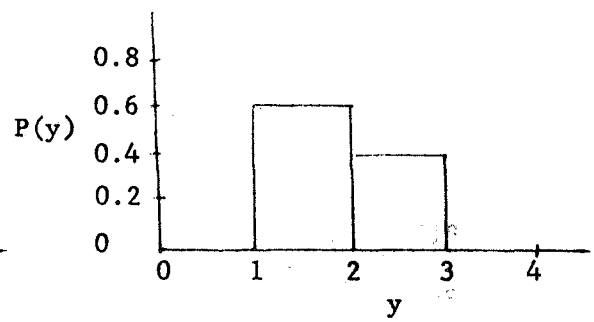
3. Sum of two equal arbitrarily distributed variables

Given that

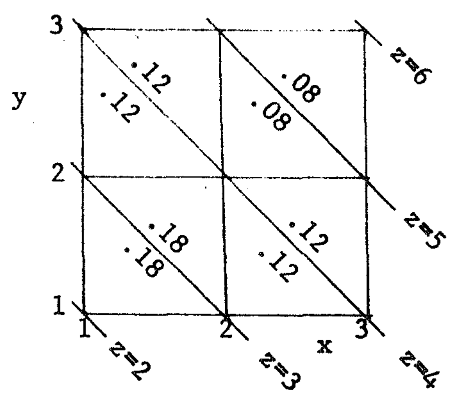
$P(1 \leq x < 2) = P(1 \leq y < 2) = 0.6$

$P(2 \leq x < 3) = P(2 \leq y < 3) = 0.4$





For $z = x + y$



$P(2 \leq z < 3) = 0.18$

$P(3 \leq z < 4) = 0.42$

$P(4 \leq z < 5) = 0.32$

$P(5 \leq z < 6) = 0.08$

$\bar{z} = 3.8$

Or, by the MBPM method

$P(2 \leq z < 4) = 0.60$

$P(4 \leq z < 6) = 0.40$

$\bar{z} = 3.8$ $\qquad \bar{x} + \bar{y} = 1.9 + 1.9 = 3.8$

4. <u>Sum of two unequal arbitrarily distributed variables</u>

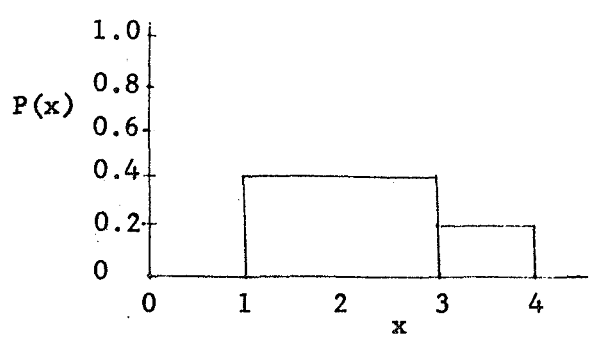
Given that

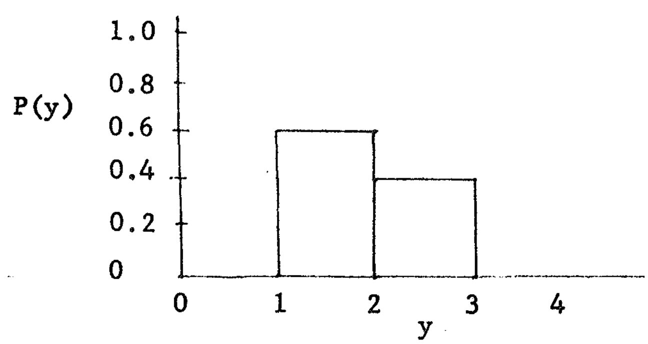
$P(1 \leq x < 3) = 0.8$

$P(3 \leq x < 4) = 0.2$

$P(1 \leq y < 2) = 0.6$

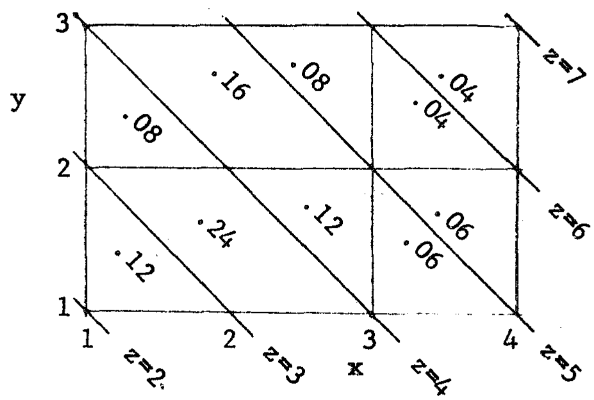
$P(2 \leq y < 3) = 0.4$





$\bar{x} = 2.3$ $\qquad \bar{y} = 1.9$

For z = x + y



$P(2 \leq z < 3) = 0.12$

$P(3 \leq z < 4) = 0.32$

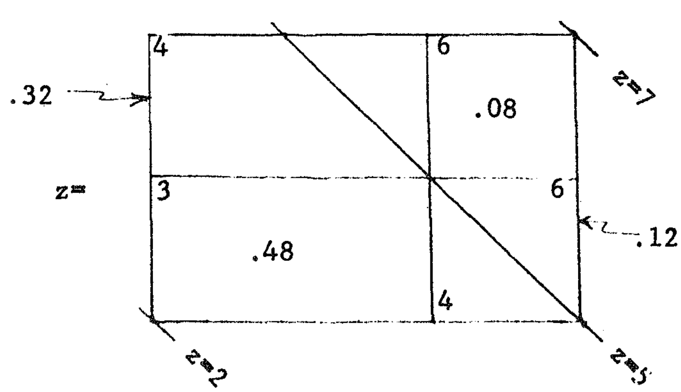
$P(4 \leq z < 5) = 0.34$

$P(5 \leq z < 6) = 0.18$

$P(6 \leq z < 7) = 0.04$

$\bar{z} = 4.2 \qquad \bar{x} + \bar{y} = 4.2$

The MBPM, prior to revision, would have calculated this as



$$P(2 \leq z < 5) = .48 + \left(\frac{5-3}{6-3}\right)(.32) + \left(\frac{5-4}{6-4}\right)(.12) = 0.75$$

$$P(5 \leq z < 7) = .08 + \left(\frac{6-5}{6-3}\right)(.32) + \left(\frac{6-5}{6-4}\right)(.12) = 0.25$$

$\bar{z} = 4.125$

The revised MBPM calculates this as

$$P(2 \leq z < 5) = .48 + \frac{3}{4}(.32) + \frac{1}{2}(.12) = 0.78$$

$$P(5 \leq z < 7) = .08 + \frac{1}{4}(.32) + \frac{1}{2}(.12) = 0.22$$

$\bar{z} = 4.05$

5. <u>Product of equal discrete variables</u>

Given the same variables as in C.1

$P(x=1) = P(y=1) = 0.25$

$P(x=2) = P(y=2) = 0.50$

$P(x=3) = P(y=3) = 0.25$

If $z = (x)(y)$, the discrete values of z will be distributed

$P(z=1) = 0.0625$

$P(z=2) = 0.25$

$P(z=3) = 0.125$

$P(z=4) = 0.25$

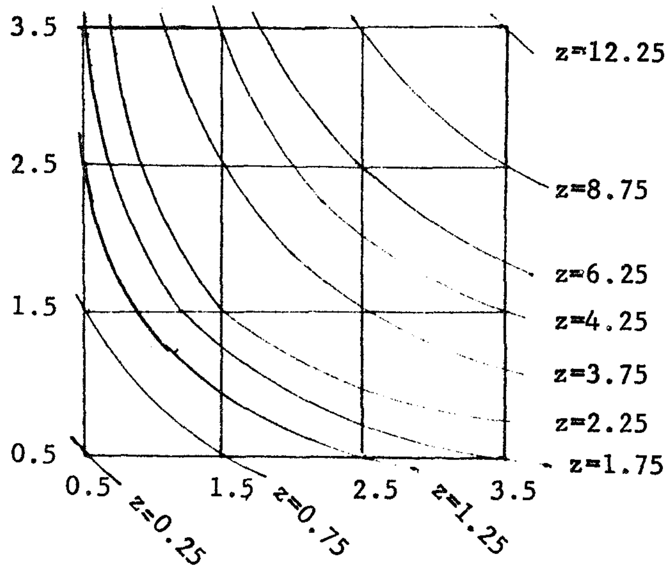
$P(z=6) = 0.25$

$P(z=9) = 0.0625$

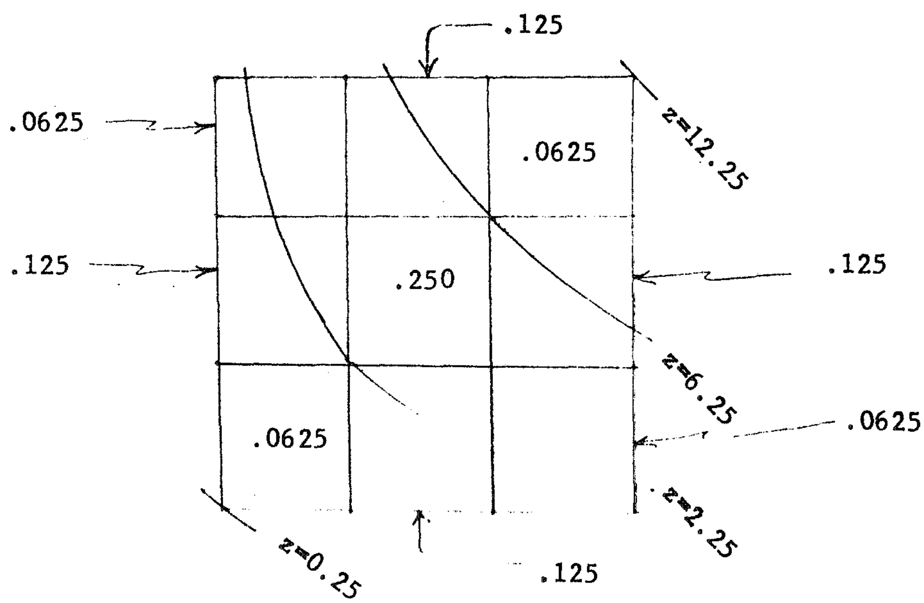
$\bar{z} = 4.00$

$(\bar{x})(\bar{y}) = (2.0)(2.0) = 4.0$

When the variables are represented by histograms, and the operation z = (x)(y) is performed, the joint probabilities are the same as for the sum, and the lines of constant z are



which the unrevised MBPM would calculate by proportion as



$$P(0.25 \leq z < 2.25) = .0625 + \left(\frac{9-3}{15-3}\right)(.125)(2) + \left(\frac{9-5}{21-5}\right)(.0625)(2)$$

$$= .0625 + .125 + .03125 = 0.21875$$

$$P(2.25 \leq z < 6.25) = .25 + \left(\frac{21-9}{21-5}\right)(.0625)(2) + \left(\frac{15-9}{15-3}\right)(.125)(2)$$

$$+ \left(\frac{25-15}{35-15}\right)(.125)(2) = .25 + .09375 + .125$$

$$+ .125 = 0.59375$$

$$P(6.25 \leq z < 12.25) = .0625 + \left(\frac{35-25}{35-15}\right)(.125)(2) = .0625 + .125$$

$$= 0.1875$$

$$\bar{z} = (.21875)\left(\frac{.25 + 2.25}{2}\right) + (.59375)\left(\frac{2.25 + 6.25}{2}\right) + (.1875)\left(\frac{6.25 + 12.2}{2}\right.$$

$$= (.21875)(1.25) + (.59375)(4.25) + (.1875)(9.25) = 4.53$$

$$(\bar{x})(\bar{y}) = (2.0)(2.0) = 4.0$$

The revised MBPM approximates the hyperbolic curves as straight lines within a joint probability segment and gives

$$P(0.25 \leq z < 2.25) = .0625 + (1)\left(\frac{1 + .4}{2}\right)(.125)(2) + (1) \quad \frac{.4 + .143}{2}$$

$$(.0625)(2) = .0625 + .175 + .03394$$

$$= 0.27144$$

$$P(2.25 \leq z < 6.25) = .25 + (1)\left(\frac{.6}{2}\right)(.125)(2) + (1)\left(\frac{.6 + .857}{2}\right)$$

$$(.0675)(2) + (1)\left(\frac{1 + .286}{2}\right)(.125)(2)$$

$$= .25 + .075 + .09106 + .16075 = 0.57681$$

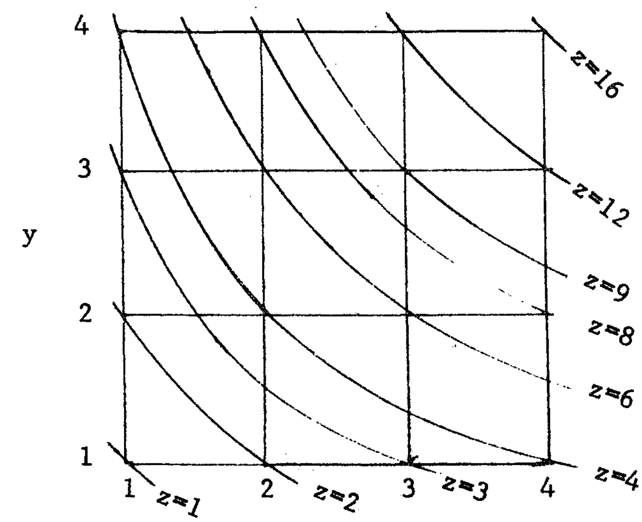
$$P(6.25 \leq z < 12.25) = .0625 + (1)\left(\frac{.714}{2}\right)(.125)(2) = .0625 + .08925$$

$$= 0.15176$$

$$\bar{z} = 4.19$$

6. <u>Product of two equal uniformly distributed variables</u>

Given the same variables as in C.2., z = (x)(y) is



by the proportion (former MBPM) method

$P(1 \leq z < 4) = .11111 + (.05556)(2) + (.02222)(2) = 0.26667$

$P(4 \leq z < 9) = .11111 + (.05556)(2) + (.08889)(2) = 0.51111$

$P(9 \leq z < 16) = .11111 + (.05556)(2) = 0.22222$

$\bar{z} = 6.77$

but $(\bar{x})(\bar{y}) = (2.5)\ (2.5) = 6.25$

by the straight-line (revised MBPM) method

$P(1 \leq z < 4) = .11111 + (1)\left(\frac{1 + 0.333}{2}\right)(.11111)(2) + (1)\ \frac{0.333}{2}\ (.11111)(2)$

$= 0.50925$

$P(9 \leq z < 16) = .11111 + (1)\ \frac{.75}{2}\ (.11111)(2) = 0.19444$

$\bar{z} = 6.48$

7. <u>Product of two normally distributed variables</u> (Ref. 10, pp. 46-56)

Given variables $X_1 = N_1\ (\mu_1, \sigma_1^2 = N_1(4, 1)$

$X_2 = N_2\ (\mu_2, \sigma_2^2 = N_2(4, 1)$

Then for $Z = \dfrac{X_1 X_2}{\sigma_1 \sigma_2 (1-\rho^2)}$ (ρ = correlation coefficient)

$$\varphi(z) = \frac{\sqrt{1-\rho^2}}{\pi} \exp\left\{ - \frac{v_1^2 - 2\rho v_1 v_2 + v_2^2}{2(1-\rho^2)} + \frac{\rho z}{(1-\rho^2)} \right\} \left[\sum_0 R_1 R_2 z) K_0(z) \right.$$

$$+ (R_1^2 + R_2^2) \frac{|z|}{2!} \sum_2 (R_1 R_2 z)\ K_1(z)$$

$$+ (R_1^4 + R_2^4) \frac{z^2}{4!} \sum_4 (R_1 R_2 z)\ K_2(z) + \ldots\ldots$$

where $v_1 = \dfrac{\mu_{x_1}}{\sigma_{x_1}}$

$v_2 = \dfrac{\mu_{x_2}}{\sigma_{x_2}}$

$$\sum_j (R_1 R_2 z) = \frac{j!}{R_1^j} \left(\frac{R_1}{R_2 z} \right)^{\frac{j}{2}} I_j \left(2\sqrt{v_1 v_2 z}\right)$$

$$K_o(z) = -\left\{ \gamma + \ln\left(\frac{1}{2} z\right) \right\} I_o(z) + \sum_{r=1} \frac{1}{(r!)^2} \left(\frac{z}{2}\right)^{2r} \left(\frac{1}{1} + \frac{1}{2} + \ldots + \frac{1}{r} \right)$$

$$= 0.57221 \ldots$$

$$K_1(z) = \left\{ \gamma + \ln\left(\frac{1}{2} z\right) \right\} I_1(z) + \frac{1}{z} - \sum_{r+1} \frac{1}{(r-1)!\,r!} \left(\frac{z}{2}\right)^{2r-1}$$

$$\cdot 1 + \frac{1}{2} + \ldots + \frac{1}{r} - \frac{1}{2r})$$

$$I_1(z) = \frac{z}{2} + \frac{z^3}{2^3 \cdot 2!} + \frac{z^5}{2^5 \cdot 2! \cdot 3!} + \frac{z^7}{2^7 \cdot 3! \cdot 4!} + \ldots\ldots$$

$$I_o(z) = 1 + \frac{z^2}{2^2 (1!)^2} + \frac{z^4}{2^4 (2!)^2} + \frac{z^6}{2^6 (3!)^2} + \ldots\ldots$$

$$R_1 = \frac{v_1 - \rho v_2}{\sqrt{1 - \rho^2}}$$

$$R_2 = \frac{v_2 - \rho v_1}{\sqrt{1 - \rho^2}}$$

$$\Phi(o) = \frac{1}{\pi} \; e^{-16} \left[\; o \; \right] = 4.143 \times 10^{-8}$$

When z is large and positive, a shorter form of K(z) may be used. (Ref. 14, p.271

For z = 4,

$$K_o(4) = \left(\frac{\pi}{8}\right)^{\frac{1}{2}} e^{-4} \left\{ 1 - \frac{1}{32} + \frac{9}{2(32)^2} - \frac{9 \cdot 25}{2 \cdot 3(32)^3} + \frac{9 \cdot 25 \cdot 49}{2 \cdot 3 \cdot 4(32)^4} - \ldots \right.$$

$$= (.01145) \left\{ .972 \right\} = .01113$$

$$K_1(4) = \left(\frac{\pi}{8}\right)^{\frac{1}{2}} e^{-4} \left\{ 1 + \frac{3}{32} - \frac{3 \cdot 5}{2(32)^2} + \frac{3 \cdot 5 \cdot 7}{2 \cdot 3(32)^3} - \frac{3 \cdot 5 \cdot 7 \cdot 9}{2 \cdot 3 \cdot 4(32)^4} + \ldots \right.$$

$$= (.01145) \left\{ 1.0875 \right\} = .01245$$

$$K_2(4) = \left(\frac{\pi}{8}\right)^{\frac{1}{2}} e^{-4} \left\{ 1 + \frac{3 \cdot 5}{32} + \frac{3 \cdot 5 \cdot 7}{2(32)^2} - \frac{3 \cdot 5 \cdot 7 \cdot 9}{3 \cdot 2(32)^3} + \frac{3 \cdot 5 \cdot 7 \cdot 9 \cdot 11}{4 \cdot 3 \cdot 2(32)^4} \ldots \right.$$

$$= (.01145) \left\{ 1.516 \right\} = .01735$$

$$\sum_0 (4 \cdot 4 \cdot 4) = 1 + (4)^3 + \frac{(4^3)^2}{4} + \frac{(4^3)^3}{4 \cdot 9} + \frac{(4^3)^4}{4 \cdot 9 \cdot 16} + \ldots\ldots$$

$$= 1 + 64 + 1024 + 7283 + 29{,}130 + 74{,}600 + 132{,}700 + 173{,}200$$

$$+ 173{,}200 + 137{,}000 + 87{,}600 + 46{,}330 + 20{,}600 + 7800 + 2546$$

$$+ 575 + 181 + 40$$

$$= 893{,}874$$

$$\sum_2 (4 \cdot 4 \cdot 4) = 1 + \frac{4^3}{3} + \frac{(4^3)^2}{(4 \cdot 3)^2} + \frac{(4^3)^3}{(5 \cdot 4 \cdot 3)(3 \cdot 2)} + \frac{(4^3)^4}{(6 \cdot 5 \cdot 4 \cdot 3)(4 \cdot 3 \cdot 2)} + \ldots\ldots$$

$$= 1 + 21 + 171 + 733 + 1942 + 3550 + 4740 + 4810 + 3850$$

$$+ 2490 + 1330 + 594 + 226 + 74 + 21 + 5$$

$$= 24{,}558$$

$$\sum_4 (4\cdot 4\cdot 4) = 1 + \frac{4^3}{5} + \frac{(4^3)^2}{(6\cdot 5)^2} + \frac{(4^3)^4}{(7\cdot 6\cdot 5)(3\cdot 2)} + \frac{(4^3)^5}{(8\cdot 7\cdot 6\cdot 5)(4\cdot 3\cdot 2)} + \ldots$$

$$= 1 + 13 + 68 + 208 + 416 + 444 + 369 + 246 + 134 + 61 + 24 + 8 + 2$$

$$= 1{,}994$$

$$\Phi(4) = 4.143 \times 10^{-8} \left[\left\{8.939 \times 10^{5}\right\} \left\{1.113 \times 10^{-2}\right\} + (64) \left\{2.456 \times 10^{4}\right\} \left\{1.245 \times 10^{-2}\right\} + (341) \left\{1.994 \times 10^{3}\right\} \left\{1.735 \times 10^{-2}\right\} + \ldots \right]$$

$$= 4.142 \times 10^{-8} \left[9.94 \times 10^{3} + 19.55 \times 10^{3} + 11.80 \times 10^{3} + \ldots \right]$$

$$= 4.143 \times 10^{-8} \left[4.129 \times 10^{4} + \ldots \right]$$

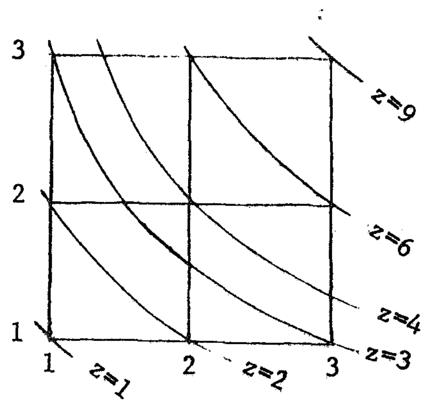
$$> 1.71 \times 10^{-3}$$

Similarly for other values of $\Phi(z)$.

8. Product of two equal arbitrarily distributed variables

Given the variables in C.3., then if $z = (x)(y)$

$(\bar{x})(\bar{y}) = 3.61$



by the proportion (former MBPM) method,

$$P(1 \leq z < 4) = .36 + \left(\frac{4-2}{6-2}\right)(.24)\ (2) = 0.60$$

$$P(4 \leq z < 9) = .16 + \left(\frac{6-4}{6-2}\right)(.24)\ (2) = 0.40$$

$\bar{z} = 4.10$

by the straight-line (revised MBPM) method

$$P(1 \leq z < 4) = .36 + (1)\left(\frac{1 + 0.33}{2}\right)(.24)(2) = 0.68$$

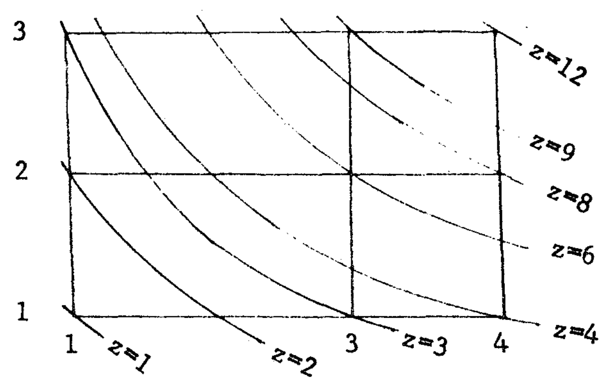
$$P(4 \leq z < 9) = .16 + (1)\left(\frac{.67}{2}\right)(.24)(2) = 0.32$$

$\bar{z} = 3.78$

9. Product of two unequal arbitrarily distributed variables

Given the variables in C.4., then if $z = (x)(y)$

$(\bar{x})(\bar{y}) \doteq (2.3)(1.9) \doteq 4.37$



by the proportion method,

$$P(1 \leq z < 6) = .48 + \left(\frac{6-2}{9-2}\right)\ (.32) + \left(\frac{6-3}{8-3}\right)\ (.12) = 0.735$$

$$P(6 \leq z < 12) = .08 + \left(\frac{9-6}{9-2}\right)\ (.32) + \left(\frac{8-6}{8-3}\right)\ (.12) = 0.265$$

$\bar{z} = 4.96$

by the straight-line method,

$$P(1 \leq z < 6) = .48 + (1)\left(\frac{3}{4}\right)(.32) + (1)\left(\frac{3}{4}\right)(.12) = 0.81$$

$$P(6 \leq z < 12) = .08 + (1)\left(\frac{1}{4}\right)(.32) + (1)\left(\frac{1}{4}\right)(.12) = 0.19$$

$$\bar{z} = 4.55$$

10. <u>End point adjustment</u>

Given that

$$P(1 \leq x < 3) = .5$$

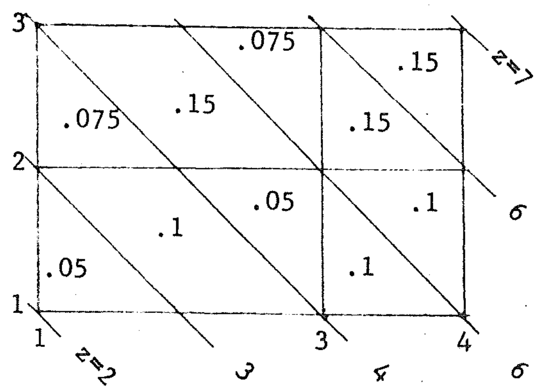
$$P(3 \leq x < 4) = .5$$

$$P(1 \leq y < 2) = .4$$

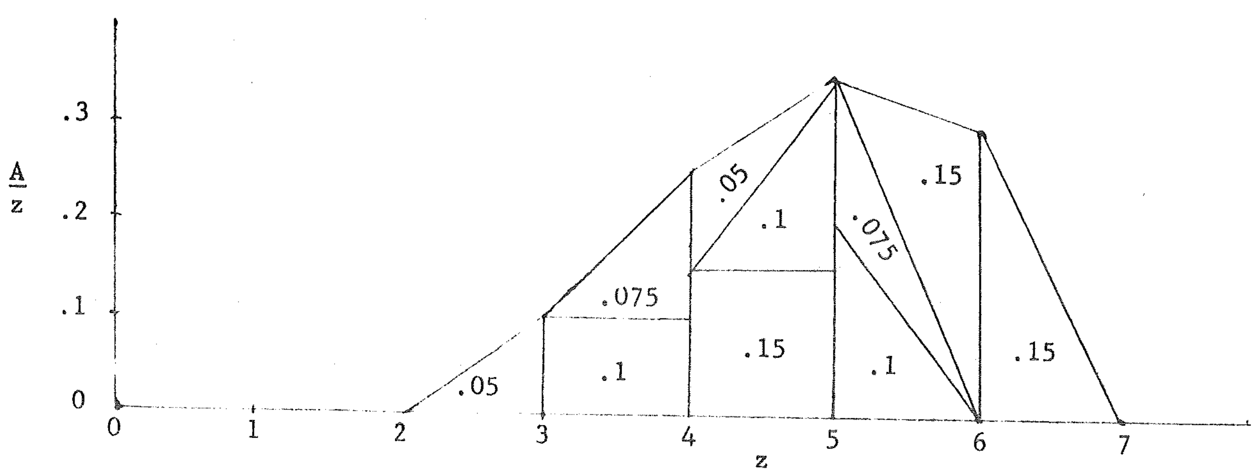
$$P(1 \leq y < 3) = .6$$

$$\bar{x} = 2.75, \ \bar{y} = 2.1$$

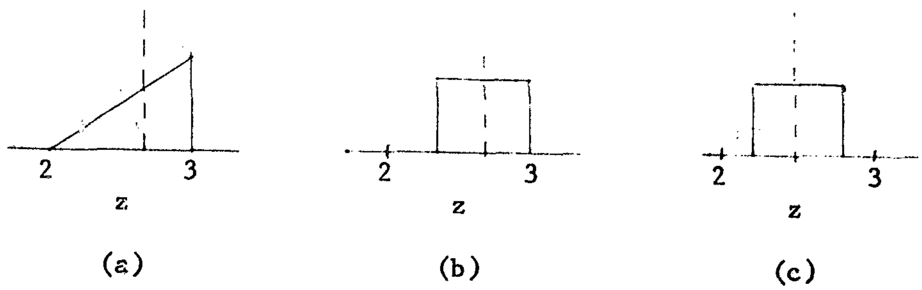
for $z = x + y$,



on an area vs. z diagram



Considering $2 \leq z < 3$ and $6 \leq z < 7$ as the end points, the lower end point may be represented by the following distributions having the same area, center of gravity and range constraints:



The dotted line represents the mean, or c.g.

(a) is accurate presentation; c.g. = 2.67; not a histogram

(b) is histogram; c.g. = 2.67; lower limit raised

(c) is histogram; c.g. = 2.5; lower limit raised; upper limit gives break before next bump

11. Skewing to correct mean

From the z-area diagram in C. 10,

Interval	Area	C.G	Moment
2<3	.05	2.67	.1335
3<4	.175	3.57	.62525
4<5	.3	4.53	1.3585
5<6	.325	5.49	1.78325
6<7	.15	6.33	.9495

$\bar{x} + \bar{y} = 4.85$

$\bar{z} = 4.90$

Reducing to two-bump,

$$\sum_{z=2}^{5} \text{Moments} = 2.12$$

$$\sum_{z=5}^{7} \text{Moments} = 2.73$$

Using the proportional method

$$\bar{z} = \left[.2 + \frac{2}{3}(.3) + \frac{1}{2}(.2)\right] (3.5) + \left[.3 + \frac{1}{3}(.3) + \frac{1}{2}(.2)\right] (6.0)$$
$$= 1.75 + 3.00 = 4.75$$

Using the straight-line method

$$\bar{z} = \left[.2 + \frac{3}{4}(.3) + \frac{1}{2}(.2)\right] (3.5) + \left[.3 + \frac{1}{4}(.3) + \frac{1}{2}(.2)\right] (6.0)$$
$$= 1.8375 + 2.85 = 4.69$$

Using the straight-line method with end-point adjustment

$$\bar{z} = \left[.2 + \frac{3}{4}(.3) + \frac{1}{2}(.2)\right]\left(\frac{2.33 + 5}{2}\right) + \left[.3 + \frac{1}{4}(.3) + \frac{1}{2}(.2)\right]\left(\frac{5 + 6.67}{2}\right)$$
$$= \left[.525\right] (3.67) + \left[.475\right] (5.83) = 1.93 - 2.77 = 4.70$$

In practice, the end-point adjustment is done on the reduced number of bumps, but is shown as above to illustrate the effect on means. The correction factor, F, required to correct the $\bar{z}$ is

$$\left(\frac{2.33 + 5F}{2}\right)(.525) + \left(\frac{5F + 6.67}{2}\right)(.475) = 4.85$$

$$0.613 + 1.313F + 1.188F + 1.5833 = 4.85$$

$$F = \frac{4.85 - 2.196}{2.5} = \frac{2.654}{2.5} = 1.06$$

$$P(2.33 \leq z < 5.3) = .525$$

$$P(5.3 \leq z < 6.67) = .475$$

$$\bar{z} = 4.85$$

12. Determination of normal distribution of input variables

On the following pages the following variables are plotted on (normal) probability paper:

(1) Logarithm of average microbial lifetime (Log v)

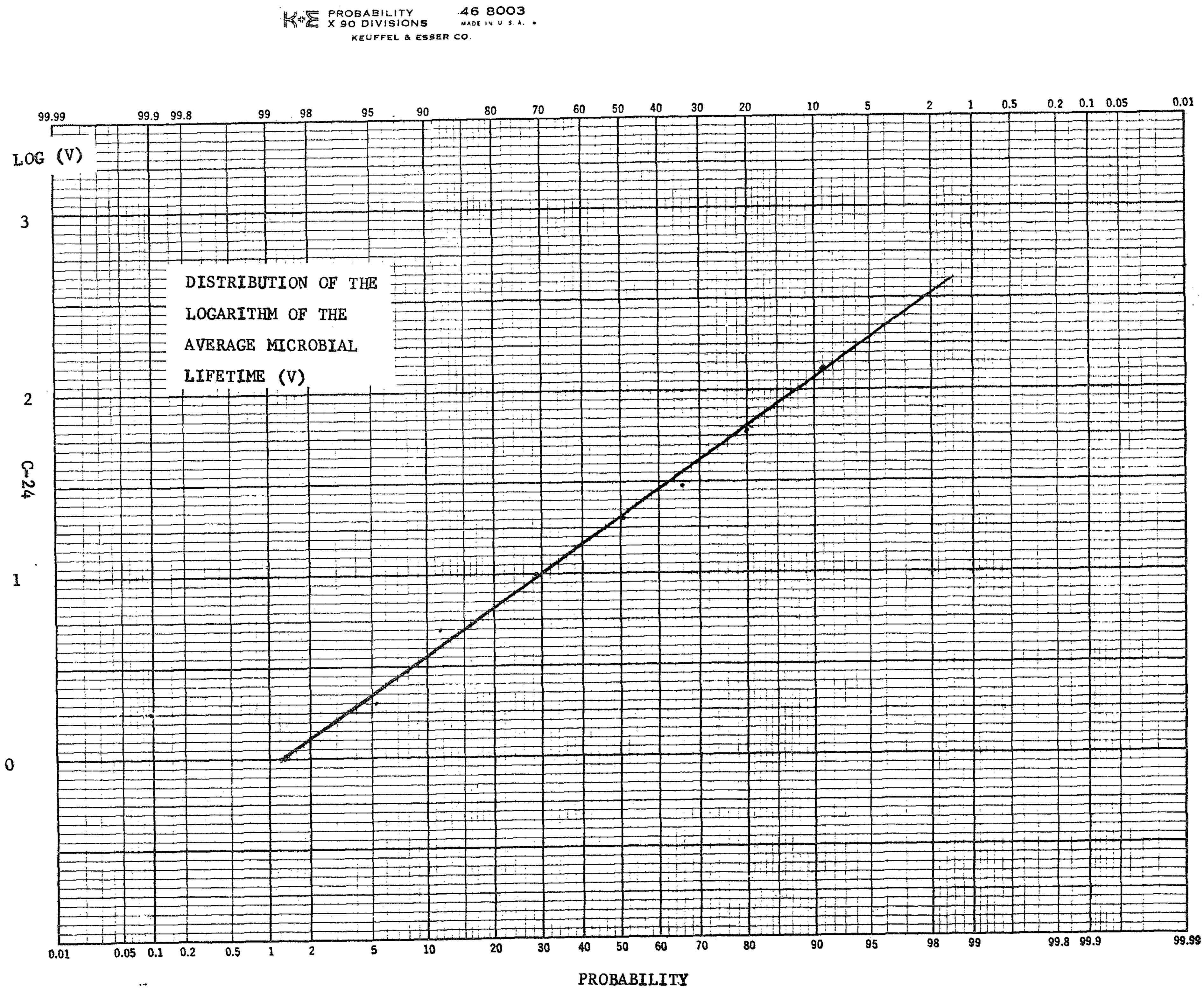
(2) Logarithm of burden on hands (Log b_t)

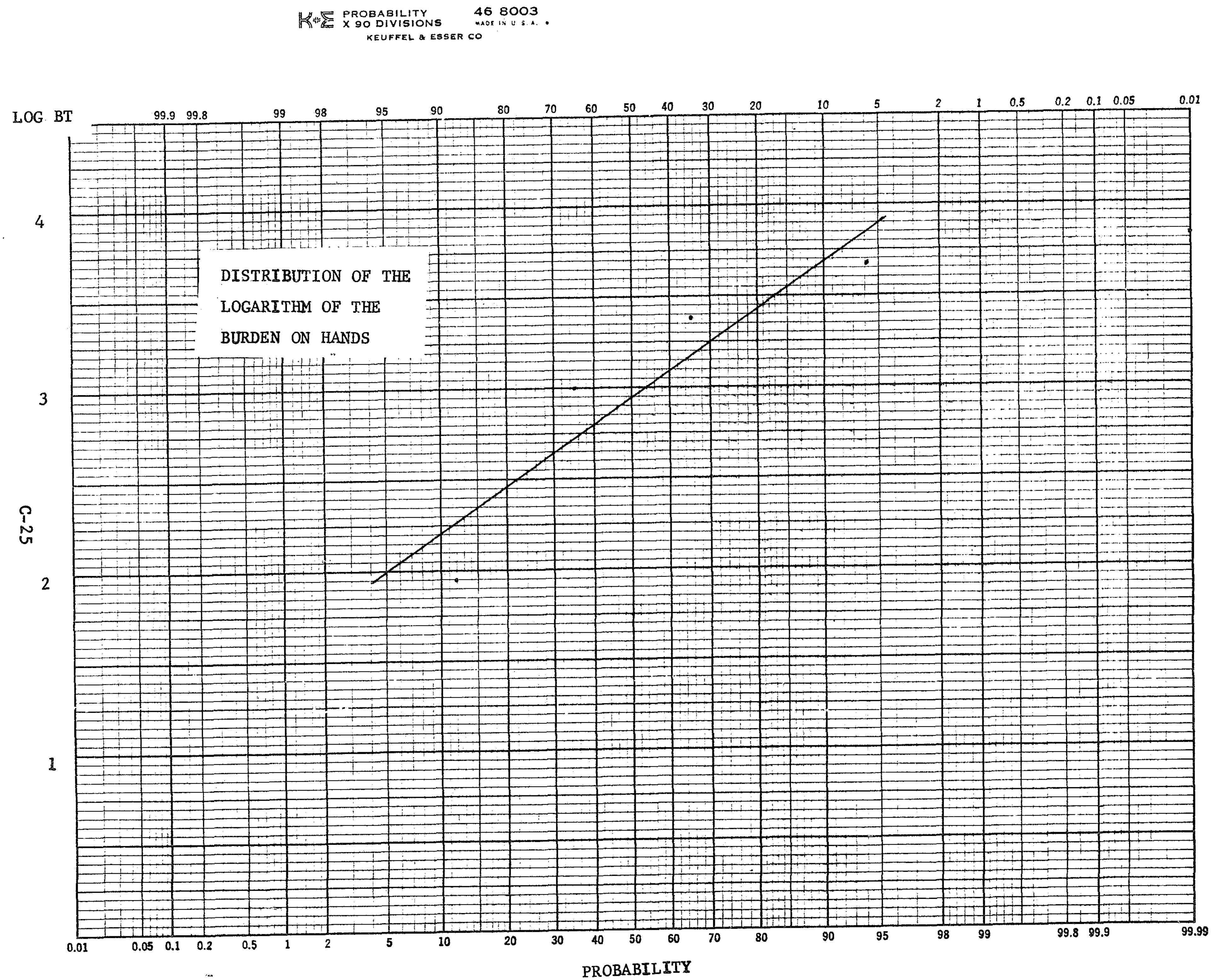
(3) Work surface retention factor (S_1)

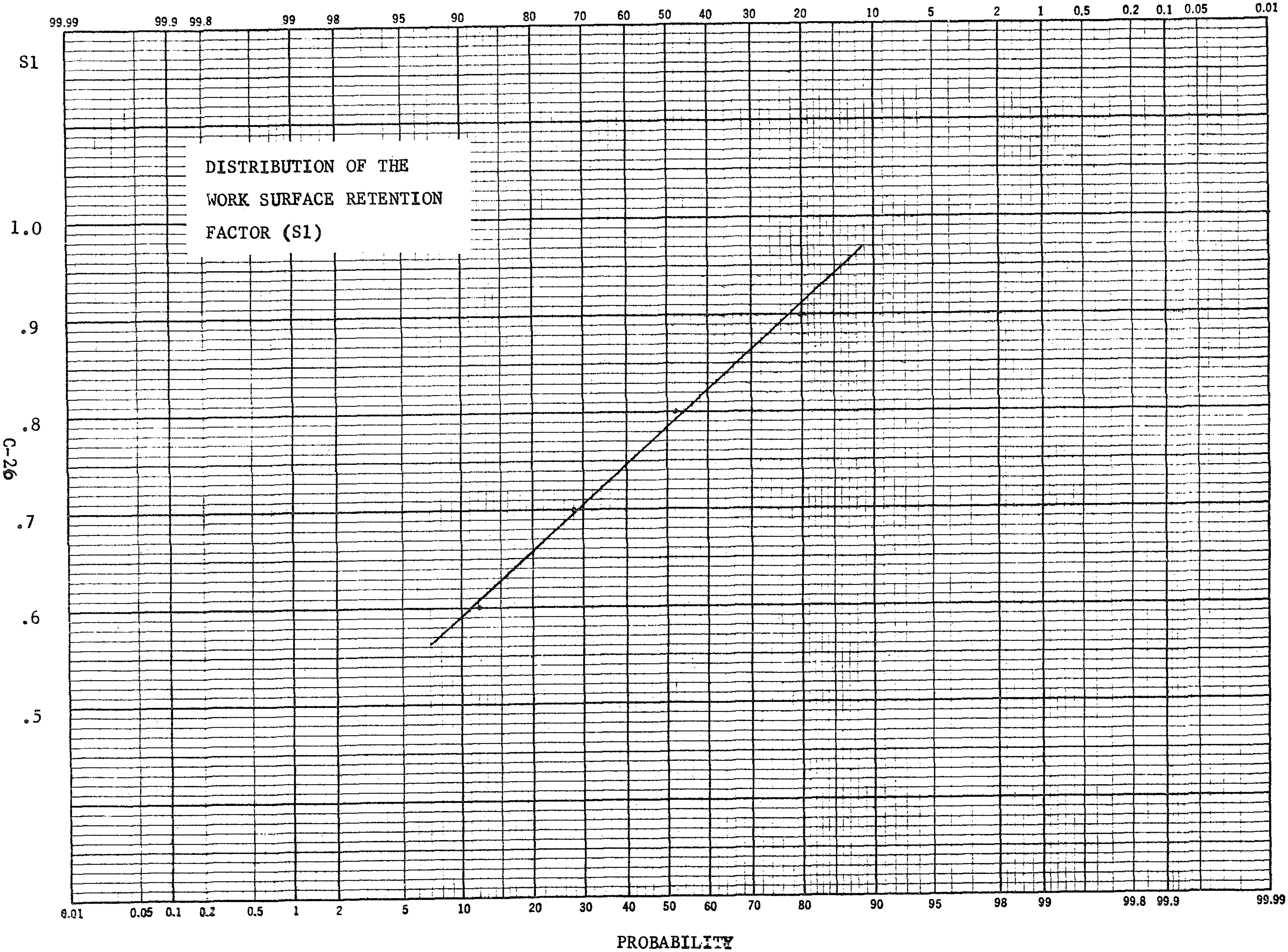
(4) Logarithm of work surface retention factor (Log S_1)

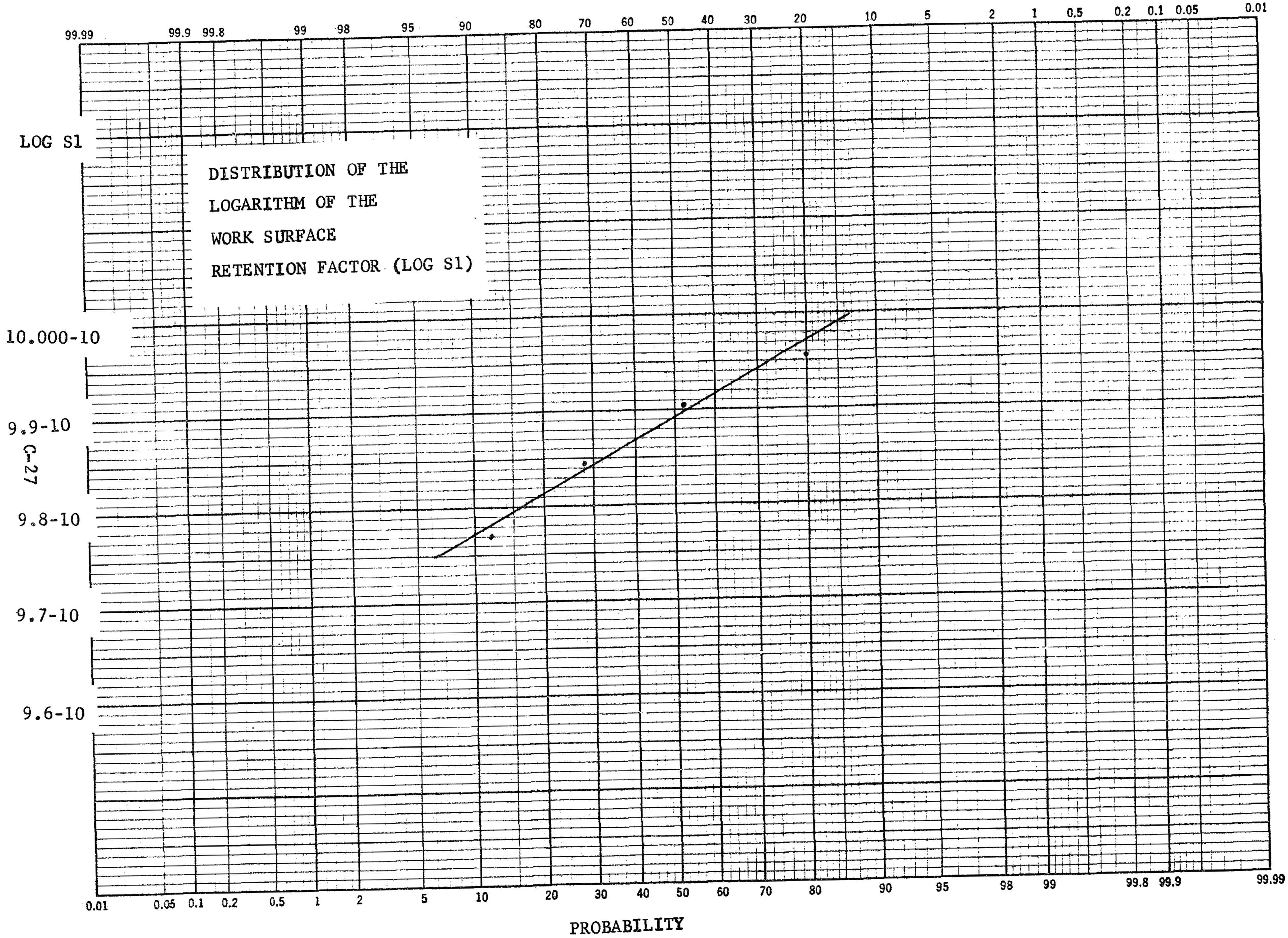
(5) Tool surface retention factor (S_2)

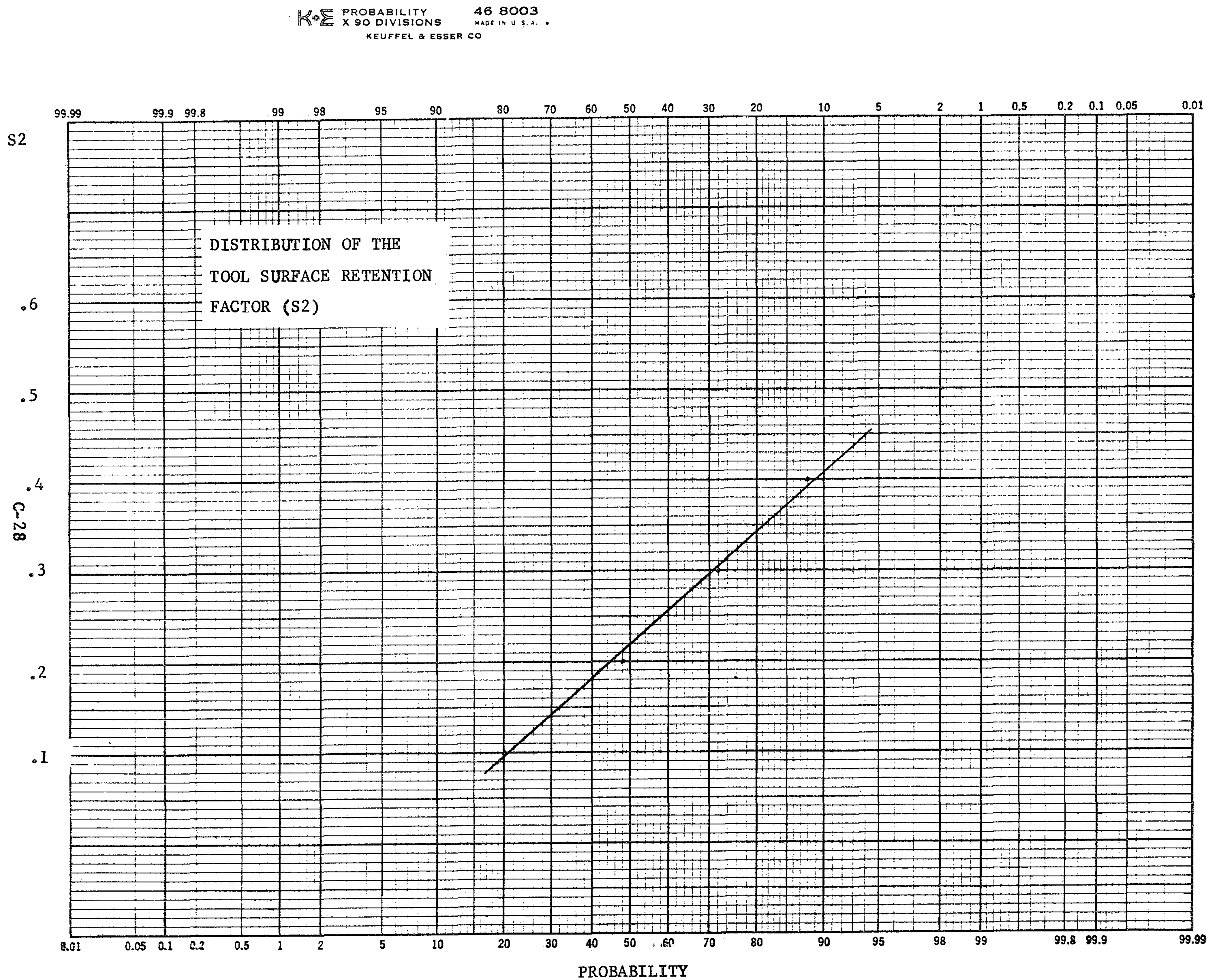
(6) Logarithm of tool surface retention factor (Log S_2)

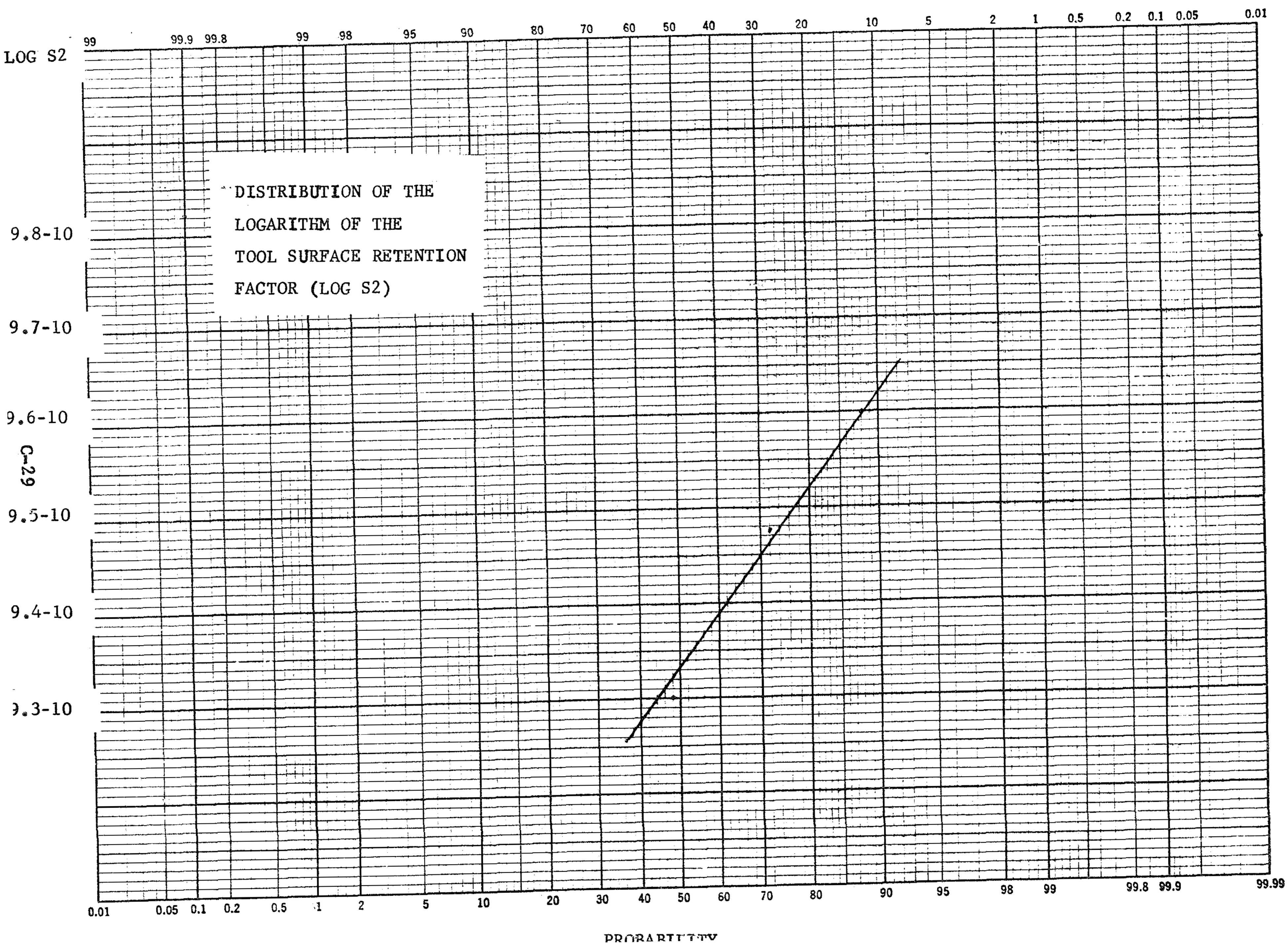

K+E PROBABILITY X 90 DIVISIONS 46 8003 MADE IN U.S.A.
KEUFFEL & ESSER CO.
LOG (V)
3
2
1
0
DISTRIBUTION OF THE LOGARITHM OF THE AVERAGE MICROBIAL LIFETIME (V)
PROBABILITY


K+E PROBABILITY X 90 DIVISIONS
46 8003
MADE IN U.S.A.
KEUFFEL & ESSER CO
LOG BT
DISTRIBUTION OF THE
LOGARITHM OF THE
BURDEN ON HANDS
4
3
2
1
99.9 99.8 99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.5 0.2 0.1 0.05 0.01
0.01 0.05 0.1 0.2 0.5 1 2 5 10 20 30 40 50 60 70 80 90 95 98 99 99.8 99.9 99.99
PROBABILITY


DISTRIBUTION OF THE
WORK SURFACE RETENTION
FACTOR (S1)
S1
1.0
.9
.8
.7
.6
.5
PROBABILITY
99.99 99.9 99.8 99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.5 0.2 0.1 0.05 0.01
0.01 0.05 0.1 0.2 0.5 1 2 5 10 20 30 40 50 60 70 80 90 95 98 99 99.8 99.9 99.99


DISTRIBUTION OF THE
LOGARITHM OF THE
WORK SURFACE
RETENTION FACTOR (LOG S1)
LOG S1
10.000-10
9.9-10
9.8-10
9.7-10
9.6-10
99.99
99.9
99.8
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.5
0.2
0.1
0.05
0.01
PROBABILITY

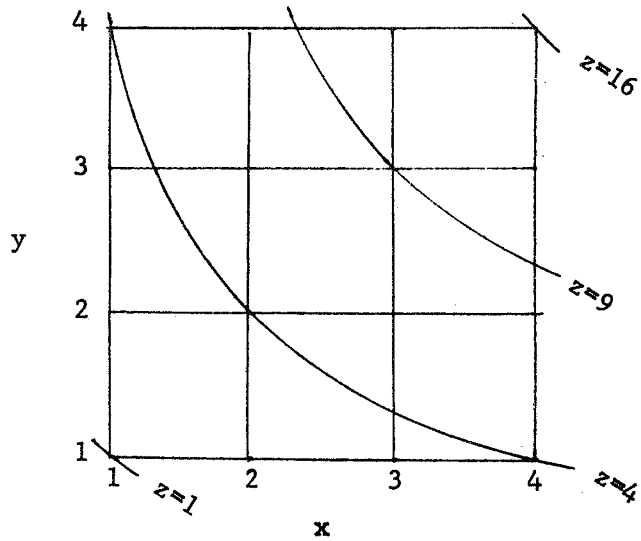

K+E PROBABILITY X 90 DIVISIONS
46 8003 MADE IN U.S.A.
KEUFFEL & ESSER CO.
DISTRIBUTION OF THE
TOOL SURFACE RETENTION
FACTOR (S2)
S2
.6
.5
.4
.3
.2
.1
99.99 99.9 99.8 99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.5 0.2 0.1 0.05 0.01
0.01 0.05 0.1 0.2 0.5 1 2 5 10 20 30 40 50 60 70 80 90 95 98 99 99.8 99.9 99.99
PROBABILITY


LOG S2
9.8-10
9.7-10
9.6-10
9.5-10
9.4-10
9.3-10
DISTRIBUTION OF THE
LOGARITHM OF THE
TOOL SURFACE RETENTION
FACTOR (LOG S2)
0.01
0.05
0.1
0.2
0.5
1
2
5
10
20
30
40
50
60
70
80
90
95
98
99
99.8
99.9
99.99
PROBABILITY

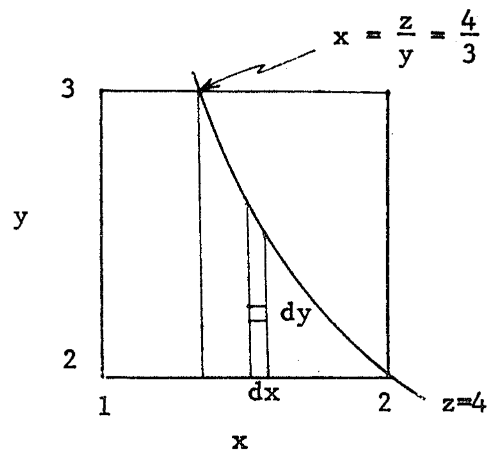
13. Other Approaches

(a) Area calculation by integration

Given the situation of C.6., i.e., $z = (x)(y)$, x and y uniform and equal distributions



$$P(1 \leq x < 2,\ 1 \leq y < 2 \mid 1 \leq z < 4) = 0.11111$$



$$P(1 \leq x < 2,\ 2 \leq y < 3 \mid z < 4) = \left\{(1)\left(\frac{1}{3}\right) + \int_{\frac{4}{3}}^{2} \int_{2}^{\frac{4}{x}} dy\, dx\right\} (.11111)$$

$$= \left\{\frac{1}{3} + \int_{\frac{4}{3}}^{2} \left(\frac{4}{x} - 2\right) dx\right\} (.11111) = \left\{\frac{1}{3} + \Big[4 \ln x - 2x\Big]_{\frac{4}{3}}^{2}\right. \quad (.11111)$$

$$= \left\{ \frac{1}{3} + \left[4(\ln 2 - \ln \frac{4}{3}) \quad - 2(2 - \frac{4}{3}) \right] \right\} (.11111)$$

$$= \left\{ \frac{1}{3} + \left[4(.6931 - .2877) - 2\left(\frac{2}{3}\right) \right] \right\} (.11111)$$

$$= \left\{ \frac{1}{3} + \left[1.6216 - 1.3333 \right] \right\} (.11111) = \left\{ .6216 \right\} (.11111) = 0.06907$$

$$P(1 \leq x < 2, 3 \leq y < 4 \mid z < 4) = \left\{ \int_1^{\frac{4}{3}} \int_3^{\frac{4}{x}} dy\, dx \right\} (.11111)$$

$$= \left[4 \ln x - 3x \right]_1^{\frac{4}{3}} (.11111) = \left[4(.2877) - 3\left(\frac{1}{3}\right) \right] (.11111)$$

$$= 0.01676$$

$$P(1 \leq z < 4) = .11111 + (.06907)(2) + (.01676)(2) = 0.28277$$

$$P(1 \leq z < 9) = 6(.11111) + 2(.11111) \left\{ \left(\frac{1}{4}\right)(1) + \int_{\frac{9}{4}}^{3} \int_3^{\frac{9}{x}} dy\, dx \right\}$$

$$= .66667 + .22222 \left\{ \frac{1}{4} + \left[9 (\ln 3 - \ln \frac{9}{4}) - 3(3 - \frac{9}{4}) \right] \right\}$$

$$= .66667 + .22222 \left\{ \frac{1}{4} + .3393 \right\} = .66667 + .13095 = 0.79762$$

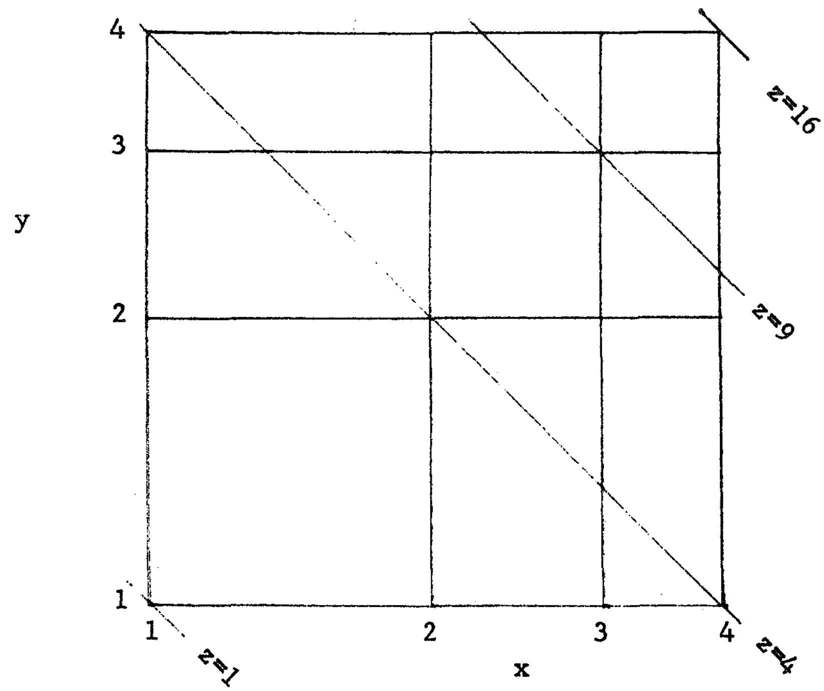
$$P(4 \leq z < 9) = .79762 - .28277 = 0.51485$$

$$P(9 \leq z < 16) = 1.00000 - .79762 = 0.20238$$

$\bar{z} = 6.58$, which is not as close to $(\bar{x})(\bar{y}) = 6.25$ as the straight-line method where $\bar{z} = 6.48$

(b) Logarithmically-based input histograms

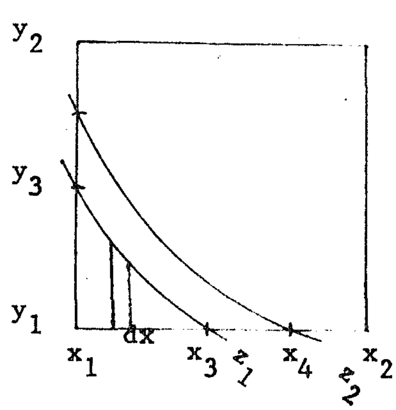
For z = (x)(y) where x and y are expressed on a logarithmic base



which apparently gives easily calculated areas, but due to non-linearity, probability is no longer proportional to area, and a method (e.g., integration) of calculating probability from the non-linear area must be used.

(c) Integrating with respect to z

An attempt was made to evaluate P(z) as z varies



$$P(z_1) = \int_{x_1}^{x_3} \left(\frac{z_1}{x} - y_1\right) dx$$

$$P(z_2) = \int_{x_1}^{x_4} \left(\frac{z_2}{x} - y_1\right) dx$$

$$P(\Delta z) = P(z_2) = P(z_1) = \int_{x_1}^{x_4} \left(\frac{z_2}{x} - y_1\right) dx - \int_{x_1}^{x_3} \left(\frac{z_1}{x} - y_1\right) dx$$

$$= z_2(\ln x_4 - \ln x_1) - y_1 (x_4 - x_1) - z_1(\ln x_3 - \ln x_1) + y_1 (x_3 - x_1)$$

$$= z_2 \ln x_4 - z_1 \ln x_3 - (z_2 - z_1) \ln x_1 - y_1 (x_4 - x_3)$$

since $z_2 - z = \Delta z$, then $y_1 x_4 - y_1 x_3 = \Delta z$

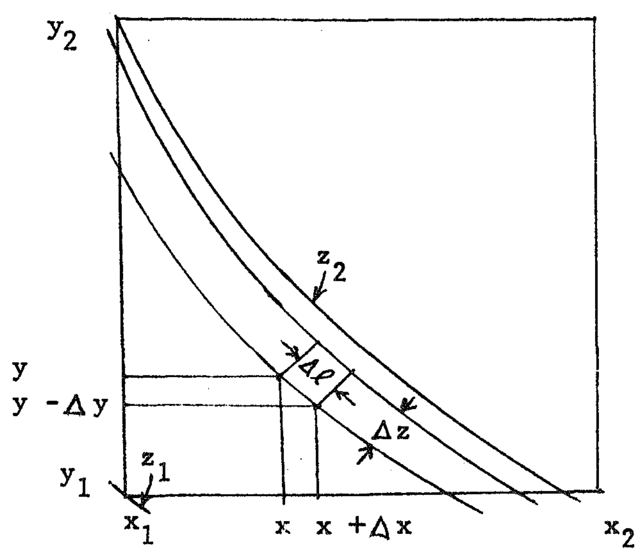
and $x_4 - x_3 = \Delta x = \frac{\Delta z}{y_1}$

$$P(\Delta z) = z_2 \ln (x_3 + \Delta x) - z_1 \ln(x_3) - \Delta z \ln x_1 - y_1 \Delta x$$

$$= (z_1 + \Delta z) \ln \left(x_3 + \frac{z}{y_1}\right) - z_1 \ln (x_3) - \Delta z \ln x_1 - \Delta z$$

$$= (z_1 + \Delta z) \ln \left(\frac{z_1}{y_1} + \frac{\Delta z}{y_1}\right) - z_1 \ln \frac{z_1}{y_1} - \Delta z \ln x_1 - \Delta z$$

Since this did not seem to be leading to a workable integral, another approach was tried



$$A = (\Delta \ell)(\Delta z)$$

$$\Delta \ell = \sqrt{(\Delta x)^2 + (\Delta y)^2}$$

$$z = xy = (x + \Delta x)(y - \Delta y) = xy + y \Delta x - x \Delta y - \Delta x \Delta y$$

$$\therefore \; y \Delta x - x \Delta y - \Delta x \Delta y = 0$$

$$\lim_{\Delta x, \Delta y \to 0} \qquad x \Delta y = y \Delta x$$

$$\Delta y = \frac{y}{x} \Delta x = \frac{z}{x^2} \Delta x$$

$$\Delta \ell = \sqrt{(\Delta x)^2 + \left(\frac{z}{x^2} \Delta x\right)^2} = \Delta x \sqrt{1 + \frac{z^2}{x^4}}$$

$$\ell = \int_{x_1}^{x_2} \sqrt{1 + \frac{z^2}{x^4}} \, dx$$

which is not readily integrated.