VARIANCES OF SOME TRUNCATED DISTRIBUTIONS FOR VARIOUS POINTS OF TRUNCATION

by

George Carlton Hayles

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

Statistics

Approved: D. R. Jensen
Chairman, Dr. D. R. Jensen

Boyd Harshbarger
Dr. B. Harshbarger

James Pickands III
Dr. J. Pickands, III

December 6, 1966

Blacksburg, Virginia

TABLE OF CONTENTS

Page

I. INTRODUCTION.................................... 6

II. MONTE CARLO PROCEDURES.......................... 14

III. THE UNIVARIATE CASE............................ 17

3.1 Procedure.................................... 17

3.2 The Standard Normal Distribution......... 19

3.3 A Pearson U-shaped Distribution.......... 23

3.4 A Bimodal Distribution................... 26

IV. THE BIVARIATE CASE............................. 30

4.1 Procedure.................................... 30

4.2 The Bivariate Normal Distribution........ 34

V. SUMMARY... 51

VI. ACKNOWLEDGEMENTS................................ 52

VII. BIBLIOGRAPHY.................................... 53

VIII. VITA.. 54

IX. APPENDIX.. 55

LIST OF TABLES

Table | Page

1. Values of the Standard Deviation of the Standard Normal Distribution Truncated at a and b($a\leq b$) Obtained by the Monte Carlo Procedure 20
2. Values of the Standard Deviation of the Standard Normal Distribution Truncated at a and b($a\leq b$) as Reported by Clark(2) 21
3. Table of Relative Differences between Corresponding Entries of Table 1 and Table 2(in%) 22
4. Values of the Variance of a Pearson U-Shaped Distribution Truncated at a and b($a\leq b$) 25
5. Values of the Variance of a Bimodal Distribution Truncated at a and b($a\leq b$) 29
6. Values of the Elements of the Covariance Matrices for Successive Regions of Truncation Expanded According to Design 1 with ρ=-.8 37
7. Values of the Elements of the Covariance Matrices for Successive Regions of Truncation Expanded According to Design 2 with ρ=-.8 38
8. Values of the Elements of the Covariance Matrices for Successive Regions of Truncation Expanded According to Design 3 with ρ=-.8 39
9. Values of the Elements of the Covariance Matrices for Successive Regions of Truncation Expanded According to Design 4 with ρ=-.8 40
10. Values of the Elements of the Covariance Matrices for Successive Regions of Truncation Expanded According to Design 1 with ρ=+.5. 41

Table | Page

11. Values of the Elements of the Covariance Matrices for Successive Regions of Truncation Expanded According to Design 2 with $\rho=+.5$ 42

12. Values of the Elements of the Covariance Matrices for Successive Regions of Truncation Expanded According to Design 3 with $\rho=+.5$ 43

13. Values of the Elements of the Covariance Matrices for Successive Regions of Truncation Expanded According to Design 4 with $\rho=+.5$ 44

LIST OF FIGURES

Figure | Page

1. Illustration showing a nested sequence of intervals used for truncating a given distribution function. 9

2. Graph of a Pearson type I U-shaped distribution ($\beta_1=0, \beta_2=1.75$) with mean equal to zero and variance equal to one. 24a

3. Graph of a bimodal distribution with mean equal to 2.5 and variance equal to 7.25 consisting of a mixture of 50% Pearson type I distribution and 50% normal distribution. 28

4. Different designs used to expand the region of truncation in the bivariate normal distribution. 33

5. Example showing how a region of truncation may be expanded by adding a "corner area". . . . 48

I. INTRODUCTION

The purpose of this study is to examine variances in the case of distributions obtained by truncating a given distribution at various points. In particular, the truncated distributions are restricted to nested increasing intervals, and the question is posed whether the variances of these distributions are monotonically increasing. The answer to this question is relevant to the use of conditional information for purposes of estimation and prediction. In order to clarify this point further, the following example is given.

Consider a case for which the expected output of a particular production process may be expressed as a function of some factors of production which represent realizations of random variables. Suppose that it further is known that variability of the product is associated with variability in the production factors, so that reduction in the variability of the latter effectively reduces the variance of the output of the process and thus improves the quality of the product.

Now suppose that we may truncate the distribution involving the factors of production in the sense that we permit only those values of the production factors which fall within certain specified intervals. If the variance of each production factor is monotonically increasing with nested increasing truncation intervals, then truncation of the distribution of the production factors may serve to reduce variability in the product. In addition to this, greater production homogeneity then could be achieved at the expense of further restrictions on the range of permissible outcomes of the production variables.

It may occur to the reader that variances in the case of successive nested truncation intervals have the monotonic property. However, this is not necessarily the case, as will be illustrated by the following example, which originally was given by Bowen(1).

Let us assume that the frequency function of X is approximated very closely by $Pr(X=-1)=1/3$, $Pr(X=0)=1/3$ and $Pr(X=k)=1/3$. Then $E(X)=(k-1)/3$ and $var(X)=2(k^2+k+1)/9$. We now exclude $X=-1$ and have $Pr(X'=0)=1/2$ and $Pr(X'=k)=1/2$, where X' denotes the random variable having the truncated distribution. Therefore, it follows that $E(X')=k/2$ and $var(X')=k^2/4$. We see immediately that $k^2/4>2(k^2+k+1)/9$ for $k\geq 9$ and, further, that the variance of X' may be made as much larger than the variance of X as we like.

A somewhat more formal statement of the problem considered here appears to be in order. To this end, let $f(X)$ be some probability density function which is of interest to and is specified by an experimenter. Let the mean of X be denoted by U_0 and let the variance of X be denoted by $var(I_0)$, where I_0 is the domain of the density function, $f(X)$. Then, by definition,

$$U_0 = \int_{-\infty}^{\infty} X f(X) dX$$

and

$$var(I_0) = \int_{-\infty}^{\infty} (X-U_0)^2 f(X) dX.$$

Now let the domain of $f(X)$ be successively truncated to the intervals I_1, I_2, ..., I_n(where n is some finite number). We will require that I_j be a subset, but not necessarily a proper subset, of I_k for any $j>k$. We define $var(I_i)$, for $i=1,2, \ldots, n$, as the variance of the random variable X when its distribution is truncated to the interval I_i, and U_i as the mean of this truncated distribution. Then

$$U_i = \frac{\int_{I_i} X f(X) dX}{\int_{I_i} f(X) dX}$$

and

$$var(I_i) = \frac{\int_{I_i} (X-U_i)^2 f(X) dX}{\int_{I_i} f(X) dX}$$

This notation is explained graphically by reference to Figure 1.

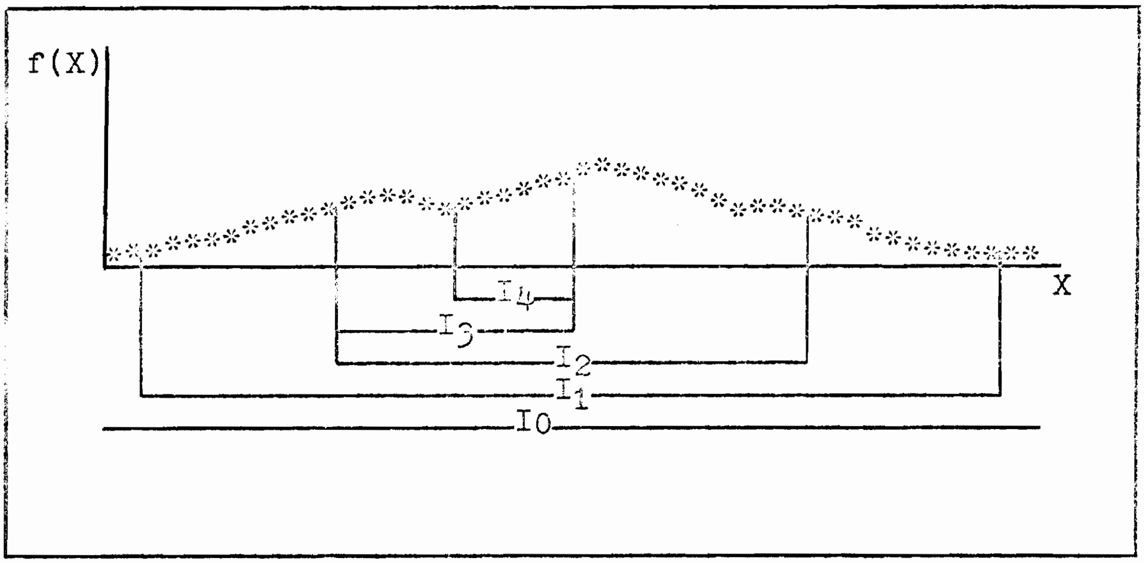


Figure 1. Illustration showing a nested sequence of intervals used for truncating a given distribution function.

For the sake of simplicity n has been set equal to four in the case illustrated in Figure 1. However, this number of intervals is sufficient to give a complete illustration of the manner in which the end points of one interval might be chosen in relation to those of another. Notice that $I_4 \subset I_3 \subset I_2 \subset I_1 \subset I_0$.

As illustrated by Figure 1, we use the convention that I_0 is the domain of the density function f(X), so that $\text{var}(I_0)$ is the variance in the case of a probability density function which has not been truncated, while I_j is the domain of the probability density function truncated to the interval I_j, where j=1, 2, 3, 4, respectively. In this thesis, empirical evidence is given to show that, for selected distributions, $\mathrm{var}(I_j) \geq \mathrm{var}(I_k)$ for $j \leq k$.

Considerable work on estimation procedures for both truncated distributions and censored samples has been done, with particular emphasis on the parameters of the original distribution. However, to the knowledge of this author, practically no work has been done to show the behavior of the values of the parameters of distributions under a sequence of truncations. For this reason, little review of literature on this problem is included here.

A table in an article presented by Clark(2) gives very clear numerical evidence of the monotonicity of variance in one particular case. He presents a table of standard deviations for the truncated standard normal distribution, including several points of truncation. It is from this table that the author has fashioned the tables for the univariate cases presented in this thesis.

From a theoretical point of view, Bowen(1) has investigated variance of truncated distributions for various truncation points. He developed several theorems and also some necessary and sufficient conditions for variances to be monotonic. Bowen was able to prove, for instance, that (i) for any truncated probability distribution, an extension of the interval of truncation chosen in such a way that it does not change the mean of the truncated distribution necessarily causes an increase in the variance of the truncated distribution; (ii) for certain other distributions which are differentiable over some known interval of truncation, if the distance between the right-hand end point of the interval and the mean is greater than the distance between the left-hand end point of the interval and the mean, the variance is monotonic for nested right-hand extensions of the interval of truncation regardless of the mean of the truncated distribution; and (iii) the variance is monotonic for nested left-hand extensions of the interval of truncation regardless of the location of the mean of the truncated distribution for certain other classes of distributions. However, there were certain other phases of the problem for which Bowen was unable to offer any type of proof of this interesting property of the variance. For instance, he was unable to prove monotonicity of variance for nested intervals in a sequence of truncations which traverses the mode of a unimodal distribution. Bowen indicated that some numerical work in this area might be most helpful.

The contents of this thesis are an extension of

Bowen's work concerning monotonicity of variance for nested intervals of truncation. The author seeks answers to questions which have practical significance but for which no theory is presently available. A Monte Carlo procedure was devised for collecting evidence regarding these questions, and the collected evidence is presented.

In this thesis, several tables are presented which provide evidence of the property of monotonic variance for nested increasing intervals of truncation in the case of univariate distributions. The Monte Carlo procedure is used to determine a table of standard deviations for the standard normal distribution with the same points of truncation reported by Clark(2). Clark's table is given intact, and it is used in comparison with the new table reported here as a check on the Monte Carlo procedure used in the present study.

Distributions other than the standard normal distribution are examined as well, namely, a Pearson U-shaped distribution and a bimodal distribution consisting of a mixture of two Pearson distributions. Graphs of the U-shaped and bimodal distributions are given.

A section is given in which dispersion for a bivariate case is examined in terms of the bivariate normal distribution. An interesting trend among the covariance matrices is observed in the data reported in that section.

The procedure used to obtain the variances presented in the tables will be discussed in some detail in a later section. First, however, a section on Monte Carlo procedures is presented which gives the reader a brief description of these techniques and shows how they have been used in the study reported here.

A separate computer program for each type of distribution was written and used to calculate the variances of the truncated distributions. FORTRAN programs and flow charts are presented in the Appendix. Explanation of the tables and procedures used to calculate the entries in the body of each table are given in each section as well as some discussion of the results presented.

II. MONTE CARLO PROCEDURES

Two alternatives appeared to be available for evaluation of the required integrals (refer again to p. 8). These two alternatives were numerical integration and Monte Carlo procedures. The numerical integration would seem to be the more accurate of the two alternatives. On the other hand, it also requires much more computer time. For example, two computer runs were made using programs written for each of these procedures. The Monte Carlo procedure gave all of the output data in 27 minutes of computer time while the numerical integration procedure gave only 1/3 of the output data in 21 minutes of computer time. The decision was made to use the Monte Carlo procedure because it was by far the more practical of the two alternatives. The Monte Carlo procedure used in this study utilized a large number of computer generated, pseudorandom numbers which may be taken to be a random sample from a specified truncated distribution. From each sample is obtained a large sample estimate of the variance of the sampled truncated distribution. Then these estimates are examined for evidence of monotonic increasing variance associated with nested increasing intervals of truncation.

By using programs which generate pseudorandom numbers from a normal distribution and a predescribed Pearson distribution available for the IBM 7040-1401 system at the Virginia Polytechnic Institute Computing Center, we may generate samples of sufficient size to give the desired accuracy. The sequence of numbers is generated one at a time by a completely predevised procedure which is, however, so devised that no significant departure from randomness may be detected by any reasonable statistical test. Numbers generated in this manner are called pseudorandom numbers. Through the use of various computer programs, sequences of such numbers are transformed into pseudorandom numbers which have any one of several probability distributions, including the normal and Pearson family mentioned earlier.

A discussion of the normal generator and how it works is given in an article by Marsaglia, MacLaren, and Bray(5). The normal generator produces pseudorandom numbers at the rate of 10,000-15,000 per second in the IBM 7040 and, according to Marsaglia, MacLaren, and Bray(5, p. 4). the method is

> "...completely accurate in the sense that in theory the procedure returns a random variable with exactly the required distribution."

To generate the random variables with the required Pearson type and mixture of Pearson type curves, a FORTRAN subroutine, which was first developed through the combined efforts of Cooper, Davis, and Dono(3) of the IBM Scientific Computational Department, is used. The subroutine was later adapted to FORTRAN IV by Donald Gale Thomas(7) for use on the IBM 7040 at the Virginia Polytechnic Institute Computing Center. The author refers the reader to Thomas' thesis for a detailed description of how the subroutine is used and for a complete FORTRAN source list and flow chart for the subroutine. The procedure generates up to 10,000 pseudorandom numbers per second from the required Pearson distribution. This subroutine is very versatile in that one is able to generate pseudorandom numbers for any type of Pearson curve or mixture of Pearson curves that is desired.

The standard normal, Pearson U-shaped and bimodal distributions are discussed in the following section. The normal generator is used in the standard normal case and Thomas'(7) subroutine is used to generate pseudorandom numbers for the last two distributions. In a later section a study of the bivariate normal distribution is presented. Again the univariate normal generator is used. A linear transformation on the univariate normal pseudorandom numbers yields the bivariate normal variables used in this study.

III. THE UNIVARIATE CASE

Procedure

In the case of the univariate distributions, a sample size of 6,000 was chosen. This number was chosen because it was found to be sufficiently large to provide accurate results as compared to Clark's table of exact values, to be discussed below, and it was not so large that it required unreasonable amounts of calculations and computer time.

A sequential type of sampling procedure was used in the program as follows. Random numbers were generated and either stored as part of the sample or discarded according to whether of not they were within the limits of truncation until the preassigned sample size was satisfied. Using the same sample size to obtain all estimates suggests that the estimates are determined with approximately equal precision.

Programs were written for further calculations involving the pseudorandom numbers generated by the procedures referred to in the preceding section. The purpose of the different FORTRAN programs used in this section is, of course, to calculate the variance of a particular truncated univariate distribution. Each program generates pseudorandom numbers from the required truncated distribution one at a time until the prescribed sample size has been achieved. Then the program calculates the required statistic and records the data along with the truncation points associated with that particular truncated distribution. Although all of these programs followed the same logical pattern, each one had to be written separately because of the different input and output requirements of each particular type of distribution. Different methods of random number generation also caused program variation. A flow chart and FORTRAN source list of the program written for the Pearson U-shaped distribution is given in the Appendix. There are no flow charts or source lists for either the standard normal or the bimodal case. They are not included because they are very similar in logic to the other programs.

The estimator $S^2 = \sum_{i=1}^{n}(X_i-\overline{X})^2/(n-1)$ is used throughout, where n=6,000 as mentioned earlier. Since the type of distribution to be sampled and consequently the mean for the distribution were known from the beginning of the experiment, one might wonder why the statistic $\sum_{i=1}^{n}(X_i-U_0)^2/n$ (where U_0 is the mean for the distribution being considered) was not used. It would seem that this estimator would give a slightly better estimate of the variance. However, we generally do not know the means of the truncated distributions for the various points of truncation.

The Standard Normal Distribution

The purpose of this section is to justify the Monte Carlo procedures used throughout this thesis. Using an exact method of calculation, Clark(2) published a table of standard deviations of the truncated standard normal distribution for various points of truncation. The purpose here is to compare with Clark's table a similar table obtained by the Monte Carlo procedure.

Table 1 contains the standard deviations for the truncated normal case, which were calculated by the Monte Carlo procedure. Truncation points are arranged in the table such that they are increasing from left to right and from bottom to top. The left hand truncation points, denoted by "a", identify columns of the table, and the right hand truncation points, denoted by "b", identify rows. The standard deviation corresponding to a pair of truncation points is located at the intersection of the column(a) and row(b) which describe the region of truncation. The range(± 3 standard deviations) and spacing(1/4 standard deviation) of the truncation points was chosen to be the same as those in Clark's table in order that the two tables might be compared.

Table 2 contains the standard deviations, as reported by Clark(2), which are associated with the same truncation points listed in Table 1. As the reader can see by comparing the two tables, the corresponding entries in Table 1

TABLE 1
VALUES OF THE STANDARD DEVIATION OF THE STANDARD NORMAL DISTRIBUTION TRUNCATED AT a AND b($a \leq b$) OBTAINED BY THE MONTE CARLO PROCEDURE

-3.00	-2.75	-2.50	-2.25	-2.00	-1.75	-1.50	-1.25	-1.00	-.75	-.50	-.25	0	.25	.50	a/b
.9912	.9818	.9752	.9556	.9408	.9155	.8766	.8276	.7886	.7491	.6992	.6392	.5968	.5525	.5008	3.00
	.9739	.9558	.9539	.9358	.9070	.8647	.8326	.7860	.7253	.6734	.6244	.5930	.5429	.4882	2.75
		.9518	.9366	.9060	.8895	.8601	.8144	.7716	.7207	.6597	.6099	.5637	.5181	.4646	2.50
			.9302	.9051	.8778	.8414	.7965	.7428	.6941	.6466	.5944	.5418	.4817	.4344	2.25
				.8662	.8489	.8091	.7725	.7335	.6682	.6207	.5591	.5035	.4445	.3853	2.00
					.8117	.7899	.7341	.6838	.6326	.5721	.5149	.4647	.3963	.3406	1.75
						.7457	.6997	.6520	.5913	.5303	.4695	.4055	.3450	.2790	1.50
							.6527	.5896	.5422	.4796	.4094	.3430	.2796	.2112	1.25
								.5382	.4802	.4183	.3531	.2840	.2141	.1417	1.00
									.4162	.3530	.2860	.2144	.1435	.0724	.75
										.2843	.2145	.1459	.0719	0	.50
											.1430	.0721	0		.25
												0			0

TABLE 2
VALUES OF THE STANDARD DEVIATION OF THE STANDARD NORMAL DISTRIBUTION
TRUNCATED AT a AND b($a \leq b$) AS REPORTED BY CLARK(2)

-3.00	-2.75	-2.50	-2.25	-2.00	-1.75	-1.50	-1.25	-1.00	-.75	-.50	-.25	0	.25	.50	a/b
.9866	.9803	.9707	.9557	.9344	.9063	.8713	.8305	.7849	.7369	.6869	.6375	.5889	.5432	.4989	3.00
	.9745	.9657	.9495	.9281	.8996	.8646	.8234	.7774	.7298	.6803	.6276	.5784	.5303	.4838	2.75
		.9546	.9394	.9176	.8891	.8535	.8118	.7605	.7155	.6639	.6120	.5610	.5106	.4610	2.50
			.9239	.9020	.8757	.8372	.7944	.7468	.6959	.6428	.5891	.5357	.4823	.4290	2.25
				.8796	.8500	.8131	.7697	.7209	.6686	.6136	.5577	.5016	.4449	.3878	2.00
					.8196	.7817	.7371	.6868	.6327	.5756	.5171	.4579	.3975	.3359	1.75
						.7425	.6966	.6446	.5885	.5291	.4679	.4055	.3414	.2792	1.50
							.6490	.5952	.5371	.4755	.4118	.3469	.2802	.2118	1.25
								.5394	.4793	.4154	.3494	.2822	.2129	.1410	1.00
									.4173	.3514	.2835	.2148	.1447	.0734	.75
										.2835	.2139	.1439	.0734	0	.50
											.1423	.0706	0		.25
												0			0

TABLE 3
TABLE OF RELATIVE DIFFERENCES BETWEEN CORRESPONDING ENTRIES
OF TABLE 1 AND TABLE 2(IN %)

-3.00	-2.75	-2.50	-2.25	-2.00	-1.75	-1.50	-1.25	-1.00	-.75	-.50	-.25	0	.25	.50	a/b
-.47	-.15	-.46	.01	-.68	-1.02	-.61	.35	-.47	-1.66	-1.79	-.27	-1.34	-1.71	-.38	3.00
	.06	1.03	-.46	-.83	-.82	-.01	-1.12	-1.11	.62	1.01	.51	-2.52	-2.38	-.91	2.75
		.29	.30	1.26	-.04	-.77	-.32	-1.46	-.73	.63	.34	-.48	-1.47	-.78	2.50
			-.68	-.34	-.24	-.50	-.26	.54	.26	-.59	-.90	-1.14	.12	-1.26	2.25
				1.52	.13	.49	-.36	-1.75	.06	-1.16	-.25	-.38	.09	.64	2.00
					.96	-1.05	.41	.44	.02	.61	-.43	-1.49	.30	-1.40	1.75
						-.43	-.45	-1.15	-.48	-.23	-.34	.00	-1.05	.07	1.50
							-.57	.94	-.95	-.86	.58	1.12	.21	.28	1.25
								.22	-.19	-.70	-1.06	-.64	-.56	-.50	1.00
									.26	-.46	-.88	.19	.83	1.36	.75
										-.28	-.28	-1.39	2.04	.00	.50
											-.49	-2.12	.00		.25
												.00			0

are very accurate in every case. In order that a comparison between the two tables may be made more easily, a table of relative differences between each corresponding entry in Table 1 and Table 2 is given as Table 3. The entries for Table 3 were computed by subtracting a particular entry in Table 1 from the corresponding entry in Table 2 and dividing by the latter. To convert the entries of Table 3 to per cent, each one was multiplied by 100. The reader will notice that the magnitude of the relative difference between every corresponding entry in the two tables is less than 3%.

One may see by observing Tables 1 and 2 that the trend of monotonicity of the standard deviation is apparent in each table. By moving up or to the left of any particular entry in either table, one observes an increase in the standard deviation. Therefore, we conclude by the evidence presented that there is definitely monotonicity of variance in the truncated standard normal case.

A Pearson U-shaped Distribution

The next distribution to be studied is another of the Pearson family of curves. This particular curve is of the Pearson type I, subclass II classification. It is well known that the Pearson type and shape of each class of Pearson curve is determined by the values of β_1 and β_2 that are chosen, where $\beta_1=(\mu_3)^2/(\mu_2)^3$, $\beta_2=(\mu_4)/(\mu_2)^2$, and μ_r is the r^{th} central moment. For the particular values of $\beta_1=0$ and $\beta_2=1.75$, the resulting Pearson type I curve is U-shaped and this curve will be studied in this section.

The U-shape of the distribution in this section is a direct contrast to the bell shaped standard normal distribution. It was for this reason that the U-shaped curve of this section was chosen. Now that we have shown evidence of the monotonic property of the variance for the bell shaped standard normal distribution, we now will present evidence of the monotonic property for a contrasting U-shaped distribution.

In order that the reader may be more familiar with the shape, range, and general outward features of the distribution, a graph of the Pearson U-shaped curve is given in Figure 2.

Table 4 gives the variances associated with different points of truncation in the Pearson U-shaped curve. The table is arranged as for the standard normal case given previously. The range on this U-shaped distribution was -1.67 to +1.67 approximately; therefore the truncation points all were restricted within this range. The distribution has mean equal to zero and variance equal to one. The spacing on the truncation points is 1/2 its standard deviation, which is somewhat larger than for the standard normal case. Therefore, the number of combinations of upper and

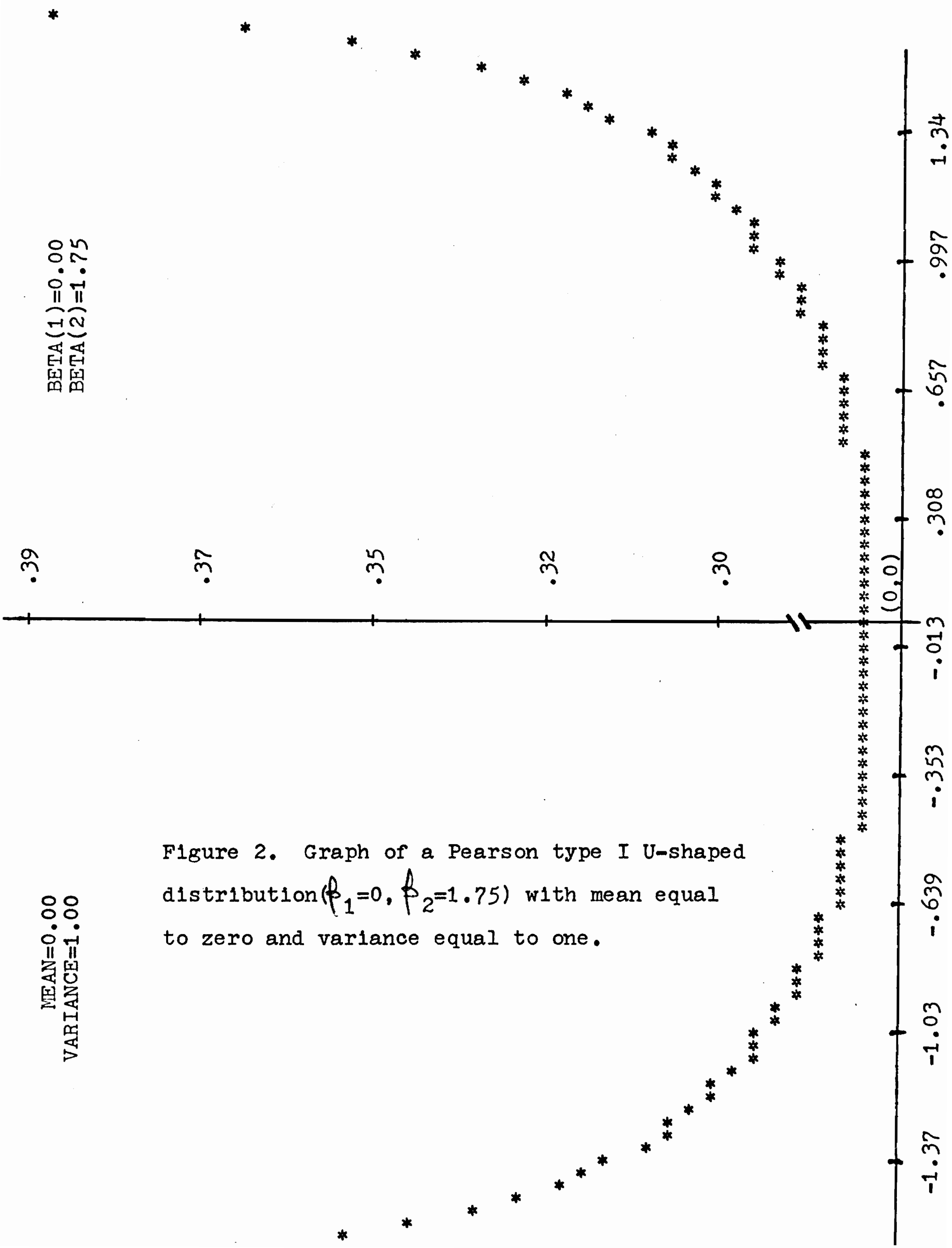


Figure 2. Graph of a Pearson type I U-shaped distribution (β_1=0, β_2=1.75) with mean equal to zero and variance equal to one.

TABLE 4

VALUES OF THE VARIANCE OF A PEARSON U-SHAPED DISTRIBUTION TRUNCATED AT a AND b($a \leq b$)

-1.67	-1.50	-1.00	-.50	0	.50	1.00	1.50	1.67	a/b
.9900	.8726	.6029	.3975	.2395	.1177	.0386	.0025	0	1.67
.8921	.7641	.5329	.3359	.1864	.0826	.0209	0		1.50
.6209	.5279	.3423	.1883	.0821	.0210	0			1.00
.4096	.3406	.1875	.0845	.0211	0				.50
.2436	.1890	.0831	.0208	0					0
.1171	.0852	.0209	0						-.50
.0385	.0210	0							-1.00
.0025	0								-1.50
0									-1.67

lower truncation points and hence the number of entries in the table is reduced. However, every combination of the upper and lower truncation points is represented in the table, and the entries there definitely show conclusive evidence of the monotonic property of the variance of a truncated Pearson U-shaped distribution.

Here, for the case of a U-shaped distribution which is in direct contrast with the bell shaped normal curve, we again see evidence of monotonicity in the variance. The reader will notice that if one starts from a particular entry in the body of Table 4 and moves to the left or in an upward direction, the successive entries are monotonically increasing. This numerical evidence thus indicates that the variance is monotonically increasing for nested increasing intervals of truncation in this U-shaped distribution.

A Bimodal Distribution

In order that the evidence presented in this section might also include other than unimodal distributions, a distribution mixture of 50% Pearson type I with mean equal to zero and variance equal to one and 50% normal with mean equal to five and variance equal to one is presented. This yields a bimodal distribution with one mode at approximately one and another mode at approximately 5.3. The distribution mixture has mean equal to 2.5 and variance equal to 7.25.

Figure 3 is a graph of this bimodal distribution. The spacing of 2.00 units between truncation points is used in this case. This particular spacing was chosen so that none of the points of truncation falls close to a mode. The truncation points were chosen in this manner so that the property of monotonicity of the variance may be studied for nested intervals of truncation traversing a mode of the distribution. This evidence complements the work done by Bowen(1) for this class of distributions. As mentioned earlier, Bowen was unable to prove that the variance was monotonic for nested intervals of truncation traversing a mode of a distribution. The evidence presented in Table 5 indicates monotonicity of the variance for nested intervals of truncation which traverse a mode.

Table 5 contains the variances for the truncated bimodal distributions. Again monotonicity of the variance is evident. Notice that for regions of truncation traversing the modes of this distribution the monotonicity of the variance is still suggested by the evidence given in Table 5. This numerical evidence is a good indication that the variance is monotonic for nested intervals of truncation in the case of this univariate bimodal distribution.

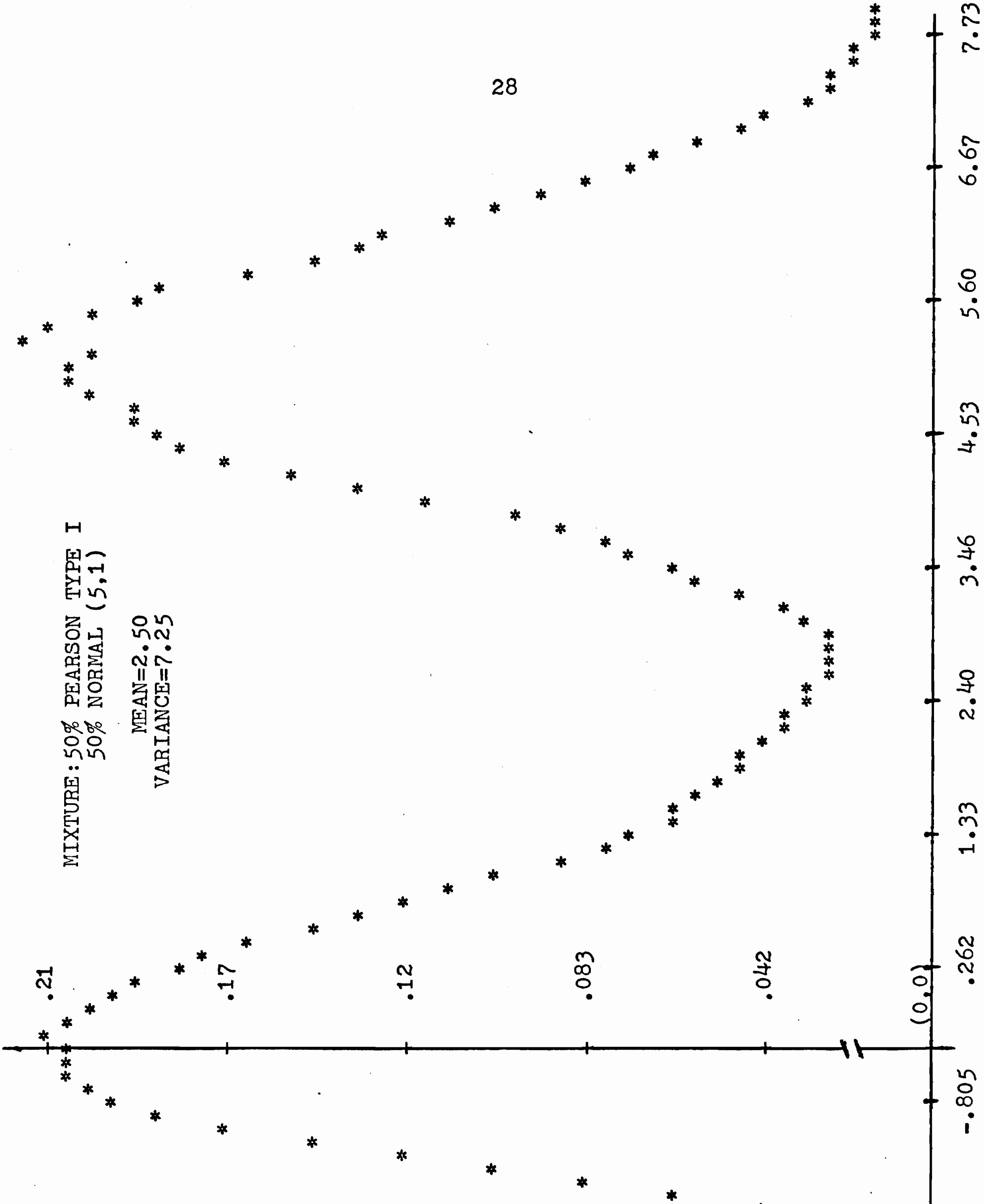


Figure 3. Graph of a bimodal distribution with mean equal to 2.5 and variance equal to 7.25 consisting of a mixture of 50% Pearson type I distribution and 50% normal distribution.

TABLE 5
VALUES OF THE VARIANCE OF A BIMODAL DISTRIBUTION TRUNCATED AT a AND b($a \leq b$)

-1.50	.50	2.50	4.50	6.50	8.00	a/b
7.2344	3.1976	.9989	.4808	.1218	0	8.00
6.7089	3.0350	.8052	.2790	0		6.50
3.4172	2.0302	.2597	0			4.50
.8195	.2741	0				2.50
.2847	0					.50
0						-1.50

IV. THE BIVARIATE CASE

Procedure

A study of the covariance matrices of a sequence of truncated bivariate normal distributions is reported in this section. The general logic and procedure of the study are similar to those of the studies conducted in the univariate case. Again the normal generator available at the Virginia Polytechnic Institute Computing Center is used. After each pair of standard normal univariate variables has been generated, a transformation on the variables must be made so that each pair together in vector form may be considered as one bivariate normal observation vector with the required mean and covariance matrix. The following is a description of the transformation in matrix notation.

Let $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ be a pair of generated standard normal random variables. Then $\underline{X}=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is distributed as $N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$

Now for any nonsingular linear transformation of the form $T\ \underline{X}=\underline{Z}$, it is known that $\underline{Z}$ has the bivariate normal distribution with mean $T\ E(\underline{X})$ and covariance matrix equal to $T\ T'=V$, say. For our present purposes it is convenient to let $T=\begin{bmatrix} t_{11} & t_{12} \\ 0 & t_{22} \end{bmatrix}$.

Because $E(\underline{X})=\underline{0}$, it follows that $E(\underline{Z})=T\ E(\underline{X})=\underline{0}$, a null vector, and thus $\underline{Z}$ is distributed as $N_2(\underline{0},E(\underline{Z}\ \underline{Z}'))$, where $E(\underline{Z}\ \underline{Z}')=E(T\ \underline{X}\ \underline{X}'\ T')=T\ E(\underline{X}\ \underline{X}')\ T'=T\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}T'=T\ T'$.

Then given a specific covariance matrix, V, we may solve for the elements of T and thus find the linear transformation which yields the required distribution. It follows that

$$t_{11}=\sqrt{v_{11}-(v_{12})^2/v_{22}}$$
$$t_{21}=0$$
$$t_{12}=v_{12}/\sqrt{v_{22}}$$
$$t_{22}=\sqrt{v_{22}}$$

where v_{11},v_{12},v_{22} are known and, since V is symmetric, $v_{21}=v_{12}$.

Then we may solve for $\underline{Z}$ since $T\ \underline{X}=\underline{Z}$ or, in matrix notation,

$$\begin{bmatrix} t_{11} & t_{12} \\ 0 & t_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

which gives

$$z_1=t_{11}x_1+t_{12}x_2$$

and

$$z_2=t_{22}x_2$$

where t_{11},t_{12},t_{22} are as described above.

After these transformations have been made, we have a bivariate normal vector with the required mean and covariance structure.

In the bivariate case, it seems reasonable that a much larger sample size should be used than in the univariate case. A sample size of 20,000 finally was selected for the bivariate normal case. A much larger sample would have been required to achieve the accuracy realized in the univariate case; however, this sample size was found to be sufficiently large to indicate behavior of the bivariate system as reported below.

The estimator $\sum_{j=1}^{n}(z_{ij}-\bar{z}_i)(z_{kj}-\bar{z}_k)/(n-1)$ for $i,k=1,2$ and where $\bar{z}_i=\sum_{j=1}^{n} z_{ij}/n$ and $\bar{z}_k=\sum_{j=1}^{n} z_{kj}/n$, which is similar to the estimator used in the univariate case, is used to get estimates of the elements of the covariance matrices. For a description of the program used for this section, see the flow chart and FORTRAN source list of the bivariate normal program in the Appendix of this thesis.

In Figure 4, different designs are given which show the various ways in which the nested regions of truncation in the bivariate case are expanded eventually to cover a large portion under the bivariate normal surface. We will require that the regions be nested rectangular regions with sides parallel to the coordinate axes. These various designs are given to enable one to visualize what area under the bivariate normal surface in being considered when reference is made to a certain number associated with one of the rectangles in Figure 4. The numbering system for

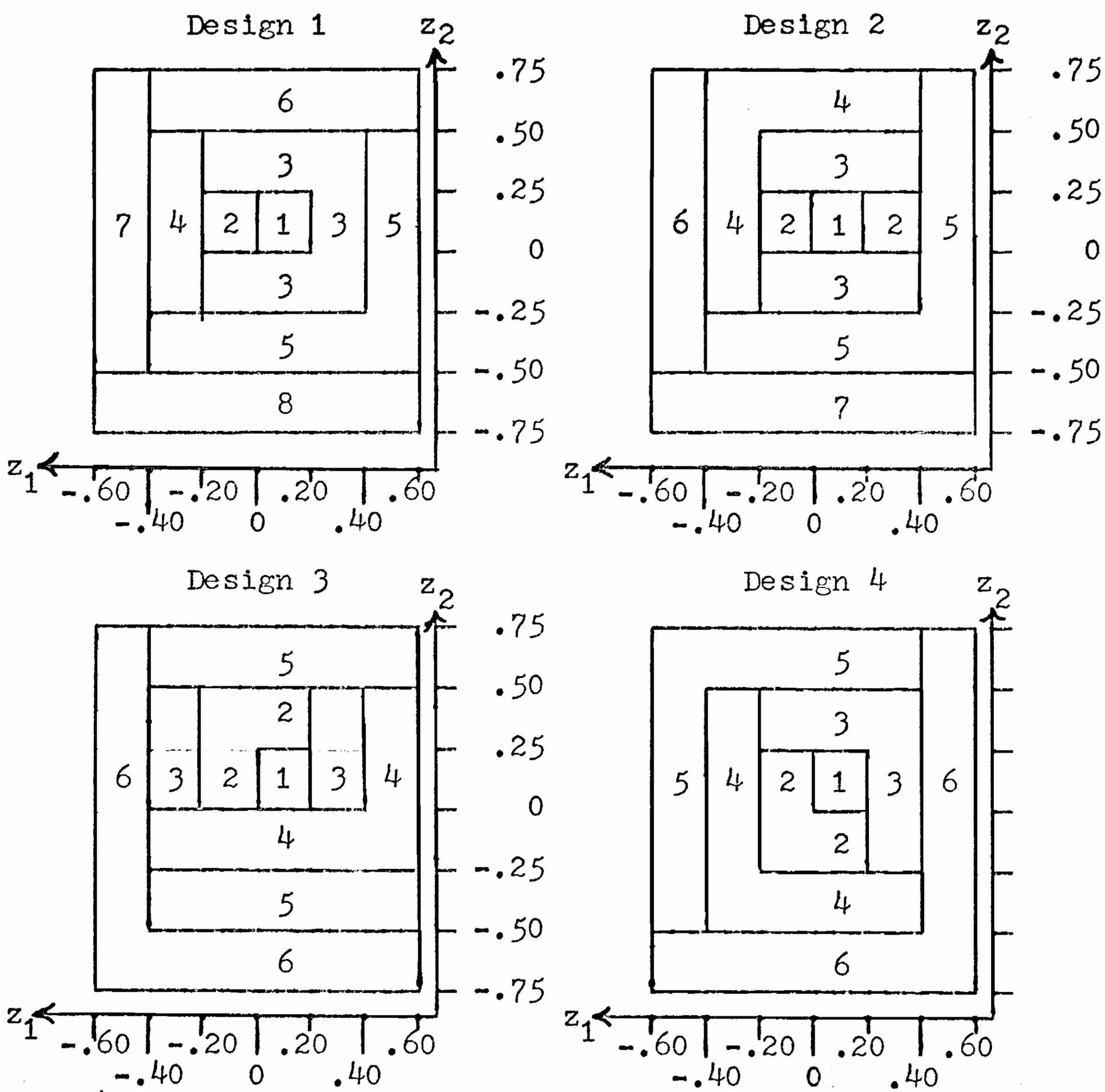


Figure 4. Different designs used to expand the region of truncation in the bivariate normal distribution.

the rectangles has the following property:

The areas with larger numbers contain all areas with smaller numbers, each of which is rectangular in shape. As an example, the area numbered as 4 is made up not only of that area labeled 4 but also the areas numbered 3, 2, and 1.

Each region of truncation is expanded horizontally along the z_1 axis by a length of $1/2\sigma_1$, where σ_1 is the standard deviation of z_1, or it is expanded vertically along the z_2 axis by a length of $1/2\sigma_2$, where σ_2 is the standard deviation of z_2, or a combination of these. As in the univariate normal case, the spacing on each variate is one-half its standard deviation.

The Bivariate Normal Distribution

The tables which follow contain the elements of the covariance matrices of various truncated bivariate normal distributions. Each set of elements contains the variances, v_{11} and v_{22}, and the covariance, v_{12}, of a truncated distribution. A number to the right of the set of elements will indicate which region of truncation is being considered. Also, at the top of each table the particular design for that table is identified.

If A and B are both 2x2 matrices (this clearly is the case for the bivariate normal distribution), then A is said to be "ordered" greater than B (denoted by A>B) if the diagonal elements of A-B and also the determinant of A-B are greater than zero. An ordering of the matrices according to the above definition will be given for each covariance matrix associated with a given region of truncation which contains another region, so that at a glance the reader may determine how each matrix compares (in order) to other matrices within any given table. Alternatively, we could consider the order only of the marginal variances of the truncated bivariate normal distributions. Since the evidence in the univariate normal case indicates that the variance is monotonic for nested intervals of truncation, we would expect the marginal variances in the bivariate normal case to be monotonic if the bivariate distribution were truncated only in one dimension. The evidence shows further that the marginal variances (the diagonal elements of the covariance matrix) were monotonic for nested regions of truncation in the respective variables. However, it appears to be useful to consider a stronger ordering in the bivariate case in which not only the variances but also the covariance are considered. Hence the ordering of positive definite matrices as defined in the preceding paragraph, which in turn implies that the marginal variances are monotonic for nested regions of truncation, but not conversely.

Table 6 gives the elements of the covariance matrices for the case $\rho=-.8$ and Design 1 of Figure 4. Notice that the line number in the table identifies each matrix, and this number corresponds to one of the areas described by the appropriate design in Figure 4. The limits on the variables z_1 and z_2, which describe the region of truncation, are given along with the ordering of the matrices. Tables 7 through 13 are tables of elements of the covariance matrices for different designs and correlation coefficients.

The covariance matrix for the bivariate normal distribution without truncation in the case of Tables 6 through 9 is

$$V = \begin{bmatrix} .16 & -.16 \\ -.16 & .25 \end{bmatrix}$$

which yields a coefficient of correlation equal to -.8. The covariance matrix for the bivariate normal distribution without truncation in the case of Tables 10 through 13 is

$$V = \begin{bmatrix} .16 & .10 \\ .10 & .25 \end{bmatrix}$$

which yields a coefficient of correlation equal to +.5.

To check the accuracy of the Monte Carlo procedure in the bivariate normal case, two runs on the computer were made in which the variances and covariance, which make up the elements of the covariance matrices given in the tables, were calculated directly by a numerical integration procedure. This procedure gives exact results, but it is costly

TABLE 6

VALUES OF THE ELEMENTS OF THE COVARIANCE MATRICES FOR SUCCESSIVE REGIONS OF TRUNCATION EXPANDED ACCORDING TO DESIGN 1 WITH $\rho = -.8$

NOTE THAT $v_{21}=v_{12}$ (V IS SYMMETRIC)

v_{11}	v_{12}	v_{22}	TRUNCATION LIMITS ON z_1 AND z_2		LINE NO.
.00319401	-.00017760	.00513278	$.00 \le z_1 \le .20$	$.00 \le z_2 \le .25$	1
.01209052	-.00065974	.00521622	$-.20 \le z_1 \le .20$	$.00 \le z_2 \le .25$	2
.02374187	-.01026568	.04272491	$-.20 \le z_1 \le .40$	$-.25 \le z_2 \le .50$	3
.04003074	-.01635724	.04339719	$-.40 \le z_1 \le .40$	$-.25 \le z_2 \le .50$	4
.05661738	-.03343158	.07314971	$-.40 \le z_1 \le .60$	$-.50 \le z_2 \le .50$	5
.05821939	-.04426287	.10314338	$-.40 \le z_1 \le .60$	$-.50 \le z_2 \le .75$	6
.07736726	-.05558026	.10484427	$-.60 \le z_1 \le .60$	$-.50 \le z_2 \le .75$	7
.08405195	-.07261732	.13979436	$-.60 \le z_1 \le .60$	$-.75 \le z_2 \le .75$	8

8 is ordered greater than 6,5,4,3,2 and 1
7 " " " " 8,5,4,3,2 and 1
6 " " " " 7,4,3,2 and 1
5 " " " " 6,4,3,2 and 1
4 " " " " 2 and 1
3 " " " " 4,2 and 1
2 " " " " 1
1 " " " " none

TABLE 7

VALUES OF THE ELEMENTS OF THE COVARIANCE MATRICES FOR SUCCESSIVE REGIONS OF TRUNCATION EXPANDED ACCORDING TO DESIGN 2 WITH $\rho=-.8$

NOTE THAT $v_{21}=v_{12}$ (V IS SYMMETRIC)

v_{11}	v_{12}	v_{22}	TRUNCATION LIMITS ON z_1 AND z_2		LINE NO.
.00325474	-.00021059	.00513854	$.00\leq z_1\leq .20$	$.00\leq z_2\leq .25$	1
.02203105	-.00137745	.00517857	$-.20\leq z_1\leq .40$	$.00\leq z_2\leq .25$	2
.02378591	-.01017345	.04291862	$-.20\leq z_1\leq .40$	$-.25\leq z_2\leq .50$	3
.04104869	-.02512187	.07075934	$-.40\leq z_1\leq .40$	$-.25\leq z_2\leq .75$	4
.05865881	-.04496950	.10503568	$-.40\leq z_1\leq .60$	$-.50\leq z_2\leq .75$	5
.07753414	-.05547413	.10489897	$-.60\leq z_1\leq .60$	$-.50\leq z_2\leq .75$	6
.08391891	-.07162510	.13851701	$-.60\leq z_1\leq .60$	$-.75\leq z_1\leq .75$	7

7 is of greater order than 5,4,3,2 and 1
6 " " " " " 7,4,3,2 and 1
5 " " " " " 6,4,3,2 and 1
4 " " " " " 3,2 and 1
3 " " " " " 1
2 " " " " " 3
1 " " " " " 2

TABLE 8

VALUES OF THE ELEMENTS OF THE COVARIANCE MATRICES FOR SUCCESSIVE REGIONS OF TRUNCATION EXPANDED ACCORDING TO DESIGN 3 WITH $\rho=-.8$

NOTE THAT $v_{21}=v_{12}$ (V IS SYMMETRIC)

v_{11}	v_{12}	v_{22}	TRUNCATION LIMITS ON z_1 AND z_2		LINE NO.
.00320280	-.00020897	.00515574	$.00\le z_1\le .20$	$.00\le z_2\le .25$	1
.01183430	-.00270492	.01981605	$-.20\le z_1\le .20$	$.00\le z_2\le .50$	2
.03578723	-.00726848	.01979045	$-.40\le z_1\le .40$	$.00\le z_2\le .50$	3
.04870672	-.01903192	.04330081	$-.40\le z_1\le .60$	$-.25\le z_2\le .50$	4
.05856048	-.04459607	.10468285	$-.40\le z_1\le .60$	$-.50\le z_2\le .75$	5
.08245746	-.07096344	.13683666	$-.60\le z_1\le .60$	$-.75\le z_2\le .75$	6

6 is ordered greater than 5,4,3,2 and 1
5 " " " " 3,2 and 1
4 " " " " 5,3,2 and 1
3 " " " " 1
2 " " " " 3 and 1
1 " " " " none

TABLE 9

VALUES OF THE ELEMENTS OF THE COVARIANCE MATRICES FOR SUCCESSIVE REGIONS OF TRUNCATION EXPANDED ACCORDING TO DESIGN 4 WITH $\rho=-.8$

NOTE THAT $v_{21}=v_{12}$ (V IS SYMMETRIC)

v_{11}	v_{12}	v_{22}	TRUNCATION LIMITS ON z_1 AND z_2		LINE NO.
.00326101	-.00020619	.00518120	$.00\leq z_1\leq .20$	$.00\leq z_2\leq .25$	1
.01218466	-.00286871	.02005045	$-.20\leq z_1\leq .20$	$-.25\leq z_2\leq .25$	2
.02383784	-.01034036	.04294308	$-.20\leq z_1\leq .40$	$-.25\leq z_2\leq .50$	3
.04327773	-.02729112	.07303275	$-.40\leq z_1\leq .40$	$-.50\leq z_2\leq .50$	4
.06211887	-.04760373	.10432240	$-.60\leq z_1\leq .40$	$-.50\leq z_2\leq .75$	5
.08294000	-.07155876	.13868141	$-.60\leq z_1\leq .60$	$-.75\leq z_2\leq .75$	6

6 is ordered greater than 5,4,3,2 and 1
5 " " " " 4,3,2 and 1
4 " " " " 3,2 and 1
3 " " " " 2 and 1
2 " " " " 1
1 " " " " none

TABLE 10
VALUES OF THE ELEMENTS OF THE COVARIANCE MATRICES FOR SUCCESSIVE REGIONS OF TRUNCATION EXPANDED ACCORDING TO DESIGN 1 WITH $\rho=+.5$
NOTE THAT $v_{21}=v_{12}$ (V IS SYMMETRIC)

v_{11}	v_{12}	v_{22}	TRUNCATION LIMITS ON z_1 AND z_2		LINE NO.
.00328826	.00012544	.00519928	$.00\le z_1\le .20$	$.00\le z_2\le .25$	1
.01280738	.00028824	.00518872	$-.20\le z_1\le .20$	$.00\le z_2\le .25$	2
.02730604	.00400513	.04367065	$-.20\le z_1\le .40$	$-.25\le z_2\le .50$	3
.04461080	.00582958	.04344860	$-.40\le z_1\le .40$	$-.25\le z_2\le .50$	4
.06302265	.01508369	.07239610	$-.40\le z_1\le .60$	$-.50\le z_2\le .50$	5
.06434505	.02062113	.10384182	$-.40\le z_1\le .60$	$-.50\le z_2\le .75$	6
.08258795	.02642927	.19469177	$-.60\le z_1\le .60$	$-.50\le z_2\le .75$	7
.08409103	.03456605	.13732998	$-.60\le z_1\le .60$	$-.75\le z_2\le .75$	8

8 is ordered greater than 6,5,4,3,2 and 1
7 " " " " 8,5,4,3,2 and 1
6 " " " " 7,5,4,3,2 and 1
5 " " " " 4,3,2 and 1
4 " " " " 2 and 1
3 " " " " 4,2 and 1
2 " " " " none
1 " " " " 2

TABLE 11
VALUES OF THE ELEMENTS OF THE COVARIANCE MATRICES FOR
SUCCESSIVE REGIONS OF TRUNCATION EXPANDED ACCORDING TO DESIGN 2 WITH $\rho=+.5$
NOTE THAT $v_{21}=v_{12}$ (V IS SYMMETRIC)

v_{11}	v_{12}	v_{22}	TRUNCATION LIMITS ON z_1 AND z_2		LINE NO.
.00328808	.00006610	.00517342	$.00\leq z_1\leq .20$	$.00\leq z_2\leq .25$	1
.02687363	.00037104	.00514057	$-.20\leq z_1\leq .40$	$.00\leq z_2\leq .25$	2
.02744215	.00371782	.04293687	$-.20\leq z_1\leq .40$	$-.25\leq z_2\leq .50$	3
.04432573	.01004045	.07054534	$-.40\leq z_1\leq .40$	$-.25\leq z_2\leq .75$	4
.06417254	.02129119	.10443606	$-.40\leq z_1\leq .60$	$-.50\leq z_2\leq .75$	5
.08360048	.02808699	.10347884	$-.60\leq z_1\leq .60$	$-.50\leq z_2\leq .75$	6
.08333572	.03620310	.13695728	$-.60\leq z_1\leq .60$	$-.75\leq z_2\leq .75$	7

7 is ordered greater than 5,4,3,2 and 1
6 " " " " 7,4,3,2 and 1
5 " " " " 6,4,3,2 and 1
4 " " " " 3,2 and 1
3 " " " " 2 and 1
2 " " " " none
1 " " " " 2

TABLE 12

VALUES OF THE ELEMENTS OF THE COVARIANCE MATRICES FOR SUCCESSIVE REGIONS OF TRUNCATION EXPANDED ACCORDING TO DESIGN 3 WITH $\rho=+.5$

NOTE THAT $v_{21}=v_{12}$ (V IS SYMMETRIC)

v_{11}	v_{12}	v_{22}	TRUNCATION LIMITS ON z_1 AND z_2		LINE NO.
.00325766	.00013318	.00517880	$.00\le z_1\le .20$	$.00\le z_2\le .25$	1
.01263596	.00073267	.01989583	$-.20\le z_1\le .20$	$.00\le z_2\le .50$	2
.04433845	.00256127	.01983734	$-.40\le z_1\le .40$	$.00\le z_2\le .50$	3
.06346914	.00899456	.04281521	$-.40\le z_1\le .60$	$-.25\le z_2\le .50$	4
.06503693	.02154120	.10584495	$-.40\le z_1\le .60$	$-.50\le z_2\le .75$	5
.08405649	.03542934	.13792921	$-.60\le z_1\le .60$	$-.75\le z_2\le .75$	6

6 is ordered greater than 5,4,3,2 and 1
5 " " " " 3,2 and 1
4 " " " " 5,3,2 and 1
3 " " " " 1
2 " " " " 3 and 1
1 " " " " none

TABLE 13
VALUES OF THE ELEMENTS OF THE COVARIANCE MATRICES FOR SUCCESSIVE REGIONS OF TRUNCATION EXPANDED ACCORDING TO DESIGN 4 WITH $\rho=+.5$
NOTE THAT $v_{21}=v_{12}$ (V IS SYMMETRIC)

v_{11}	v_{12}	v_{22}	TRUNCATION LIMITS ON z_1 AND z_2		LINE NO.
.00329565	.00012640	.00515649	$.00\le z_1\le .20$	$.00\le z_2\le .25$	1
.01269666	.00083982	.02015157	$-.20\le z_1\le .20$	$-.25\le z_2\le .25$	2
.02710635	.00369031	.04371975	$-.20\le z_1\le .40$	$-.25\le z_2\le .50$	3
.04574200	.01025831	.07277646	$-.40\le z_1\le .40$	$-.50\le z_2\le .50$	4
.06189500	.01930712	.10415227	$-.60\le z_1\le .40$	$-.50\le z_2\le .75$	5
.08345857	.03547317	.13863448	$-.60\le z_1\le .60$	$-.75\le z_2\le .75$	6

6 is ordered greater than 5,4,3,2 and 1
5 " " " " 4,3,2 and 1
4 " " " " 3,2 and 1
3 " " " " 2 and 1
2 " " " " 1
1 " " " " none

in computer time, as might be expected. For this reason the numerical integration procedure was not used extensively in this study. However, it is worthwhile to note these results and make comparisons between them and the results obtained by the Monte Carlo procedure. Two such checks were made for comparison with the Monte Carlo results. The regions of truncation for the first and second checks, respectively, are described by the following limits on the variables z_1 and z_2:

Check #1	Check #2
$.00 \le z_1 \le .20$	$-.40 \le z_1 \le .60$
$.00 \le z_2 \le .25$	$-.25 \le z_2 \le .50$

Each check was made with correlation coefficient equal to -.8. The results by the numerical integration procedure were as follows :

Check	v_{11}	v_{12}	v_{22}
1	.00325295	-.00001387	.00520086
2	.04966832	-.00105016	.04687998

The results for the same regions of truncation obtained by the Monte Carlo procedure were as follows:

Check	v_{11}	v_{12}	v_{22}
1	.00320280	-.00020897	.00515574
2	.04870672	-.01903192	.04330081

To help facilitate comparison of the results from the two procedures, a listing of the absolute error between the two follows:

Check	v_{11}	v_{12}	v_{22}
1	.00005015	.00019510	.00004512
2	.00096160	.01798176	.00357917

The absolute differences in the results are not particularly small when compared with the results attained in the univariate normal case but, in order to achieve equivalent accuracy in the bivariate normal case, it seems logical to use a sample size which is on the order of the square of the sample size used in the univariate normal case. The resulting sample size of 36,000,000 would be unreasonably large. In view of this fact, perhaps the accuracy of the results is acceptable if we consider the limitation on the sample size used. However, the sample size appears to have been sufficiently large to reflect consistent trends in the covariance matrices of the truncated distributions.

Now that some justification has been given the procedure used, we now will discuss the evidence given in the tables, point out certain trends, and draw conclusions from the evidence presented.

We see in this bivariate case that a successive nested increase in the region of truncation does not necessarily increase the order of the covariance matrix. In all of the different tables presented, the regions of truncation are nested increasing regions and, by observing the ordering of the successive matrices, we see that an increase in the area of the region of truncation sometimes does and sometimes does not cause the covariance matrix of the region after expansion to be of greater order than that of the region considered before expansion. Therefore, in contrast to conclusions made in the univariate case, no categorical conclusions can be made in the bivariate normal case. However, one important trend should be pointed out at this time. If the reader will use the appropriate design for the particular table being considered and observe in what manner the region of truncation is being expanded, it will aid recognition of the following trend in the order of the successive covariance matrices. If the region of truncation is expanded by extending the limits in one or both directions on both variables z_1 and z_2 simultaneously, the resulting covariance matrix in all cases is of greater order than the covariance matrix of the region considered previously. In other words, if one observes the design, one will notice that if a "corner area" is added to the region of truncation, as in Figure 5, then the covariance matrix for area 2 (which, we recall, is made up of the areas labeled 1 and 2) is of greater order than the covariance matrix for area 1. This trend may be noted throughout the tables given for the bivariate normal distribution.

In order that we may observe this trend, consider the following specific cases in which "corner areas" are added in expanding the limits of truncation. In particular,

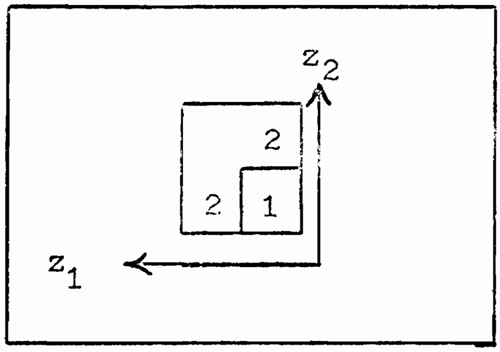


Figure 5. Example showing how a region of truncation may be expanded by adding a "corner area".

consider Tables 9 and 13. Notice that these two tables both refer to design 4 and ρ =-.8 and ρ =+.5 for Table 9 and Table 13, respectively. Design 4 is expanded in the manner illustrated by Figure 5. Observe that each successive line, starting from the top of the table and moving down, contains a covariance matrix for successive "corner area" expansions of the limits of truncation. Notice also that the matrices represented in each successive line are of increasing order when read from top to bottom. One may check the order of two matrices (say line 1 and line 2) by comparing the diagonal elements of the two matrices and observing the sign of the determinant of the difference matrix (matrix 2 minus matrix 1). In this example, the diagonal elements of matrix 2 are larger than those of matrix 1 and the determinant of the difference matrix is greater than zero, which can be verified by a few simple calculations. Thus, by the definition of ordering of matrices, matrix 2 is of greater order than matrix 1. One may check any pair of matrices in Tables 9 and 13 by the procedure given above and verify that the order of the matrices is increasing when the region of truncation is expanded by adding such "corner areas".

Other specific cases in which "corner areas" were added appear in line 5 as compared to line 4 and line 4 as compared to line 3 in Tables 7 and 11. Further examples are line 2 as compared to line 1, line 4 as compared to line 3, and line 6 as compared to line 5 in Tables 8 and 12. All of these cases bear evidence that when a "corner area" is used to expand the truncation limits, the covariance matrix of the expanded region is of greater order than the covariance matrix before expansion. Of course, all of these remarks must be taken in the context of the increments by which, in this study, any region is expanded in the direction of z_1 $(1/2\sigma_1)$ and z_2 $(1/2\sigma_2)$.

When the region of truncation is expanded on three sides (which includes a "corner area" extension), for example, in line 3 as compared to line 2 of Tables 6 and 10 and line 4 as compared to line 3 in Tables 7 and 11, the matrices are of increasing order. This result is anticipated in light of the remarks in the preceding paragraph.

Areas for further work in the bivariate and multivariate cases are almost unlimited. One could conduct a study similar to the one done here for other bivariate or multivariate distributions. The multivariate normal distribution might be a likely candidate. There is still much work to be done in the bivariate normal case. One might consider other values of the correlation coefficient. Perhaps a bivariate normal distribution with a smaller negative correlation might shed some new light on the subject.

V. SUMMARY

The univariate distributions studied in this thesis are the standard normal distribution, a Pearson type I, U-shaped distribution (β_1=0.0 and β_2=1.75), and a bimodal distribution given by mixing 50% Pearson type I distribution (β_1=.5 and β_2=3.4) and 50% normal distribution with mean equal to five and variance equal to one. We have seen that there is a definite monotonic trend in the variances of truncated distributions in the case of nested intervals of truncation. The evidence is given in Table 1 for the standard normal case, Table 4 for the case of the Pearson U-shaped distribution, and Table 5 for the bimodal case. The trend can be seen clearly by observing these tables.

As for the bivariate case, the property of monotonicity of variance is not always evident; however, specific cases were found for which (depending on how the region of truncation was expanded) evidence of the monotonic property is suggested in the sense that the positive definite covariance matrices are ordered. However, the evidence in Tables 6 through 13 indicates that the marginal variances are monotonically increasing for nested increasing regions of truncation in the respective variables regardless of the manner in which the region of truncation is expanded.

VI. ACKNOWLEDGEMENTS

The author wishes to express his appreciation to the following:

Dr. Donald R. Jensen for the time and effort spent in providing suggestions and criticisms of this thesis.

Dr. Clyde Y. Kramer for his reading and corrections of the manuscript.

Mr. Philip N. Bergstresser for his invaluable assistance in providing information on computer technology.

The Department of Health, Education, and Welfare for the fellowship which has made my graduate work at Virginia Polytechnic Institute possible.

VII. BIBLIOGRAPHY

(1) Bowen, Jacob Van, Jr., Some Properties of Conditional Distributions of A Special Type. Virginia Polytechnic Institute: Masters Thesis. 1966.

(2) Clark, F. E., "Truncation to Meet Requirements on Means", Journal of the American Statistical Association, Vol. 52 (December 1957), pp. 527-536.

(3) Cooper, J. D., Davis, S. A., and Dono, N. R., "Pearson Universal Random Distribution Generator (PURGE)". IBM Share Programs Distributions. New York: G2 IBM 003. 1963.

(4) Hammersley, J. M., and Handscomb, D. C., Monte Carlo Methods. New York: John Wiley and Sons. 1965.

(5) Marsaglia, G., MacLaren, M. D., and Bray, T. A., "A Fast Procedure for Generating Normal Random Variables", Communications of the ACM, Vol. 7, (January, 1964), pp. 4-10.

(6) Marshall, A. W., Introductory Note, Symposium on Monte Carlo Methods, ed. Meyer, H. A., New York: John Wiley and Sons, (1956), pp. 5-9.

(7) Thomas, Donald Gale, Computer Methods for Generating Pseudo-Random Numbers from Pearson Distributions And Mixtures of Pearson and Uniform Distributions. Virginia Polytechnic Institute: Masters Thesis. 1966.

VIII. VITA

George Carlton Hayles was born on September 12, 1942, in Water Valley, Mississippi. He attended elementary and secondary schools in Water Valley and was graduated from Water Valley High School in May, 1960.

In the fall of 1960, he entered Northwest Junior College in Senatobia, Mississippi and finished junior college in May, 1962 with the degree of Associate in Science in mathematics.

He enrolled at The University of Mississippi in September, 1962 and in July, 1964, he received the degree of Bachelor of Science in mathematics.

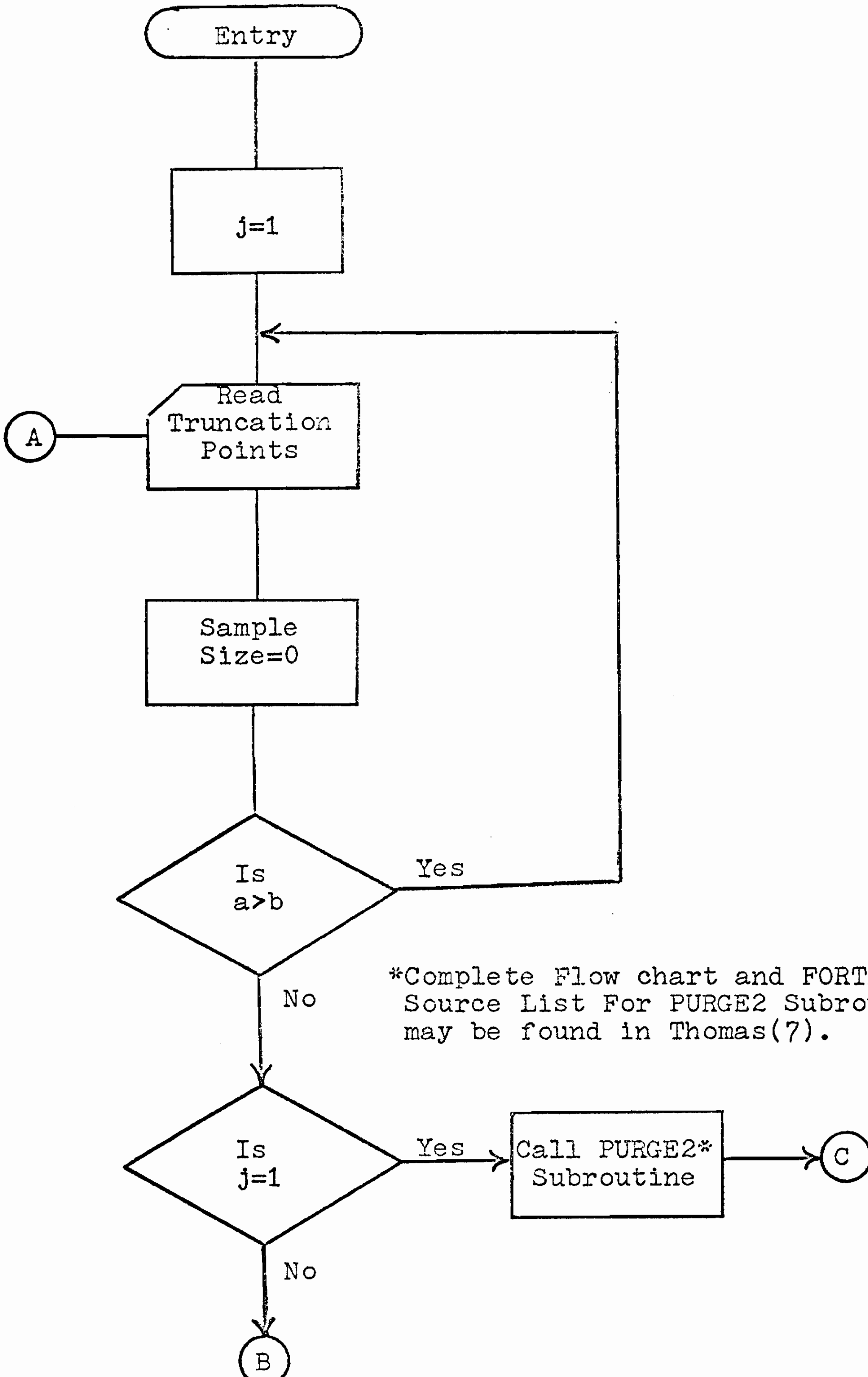
He attended The University of Mississippi for one semester in the fall of 1964 and did graduate work in mathematics.

In September, 1965, he began graduate work in statistics at Virginia Polytechnic Institute. He was married to Miss Susan Ann Brown on September 25, 1966.

George Carlton Hayles

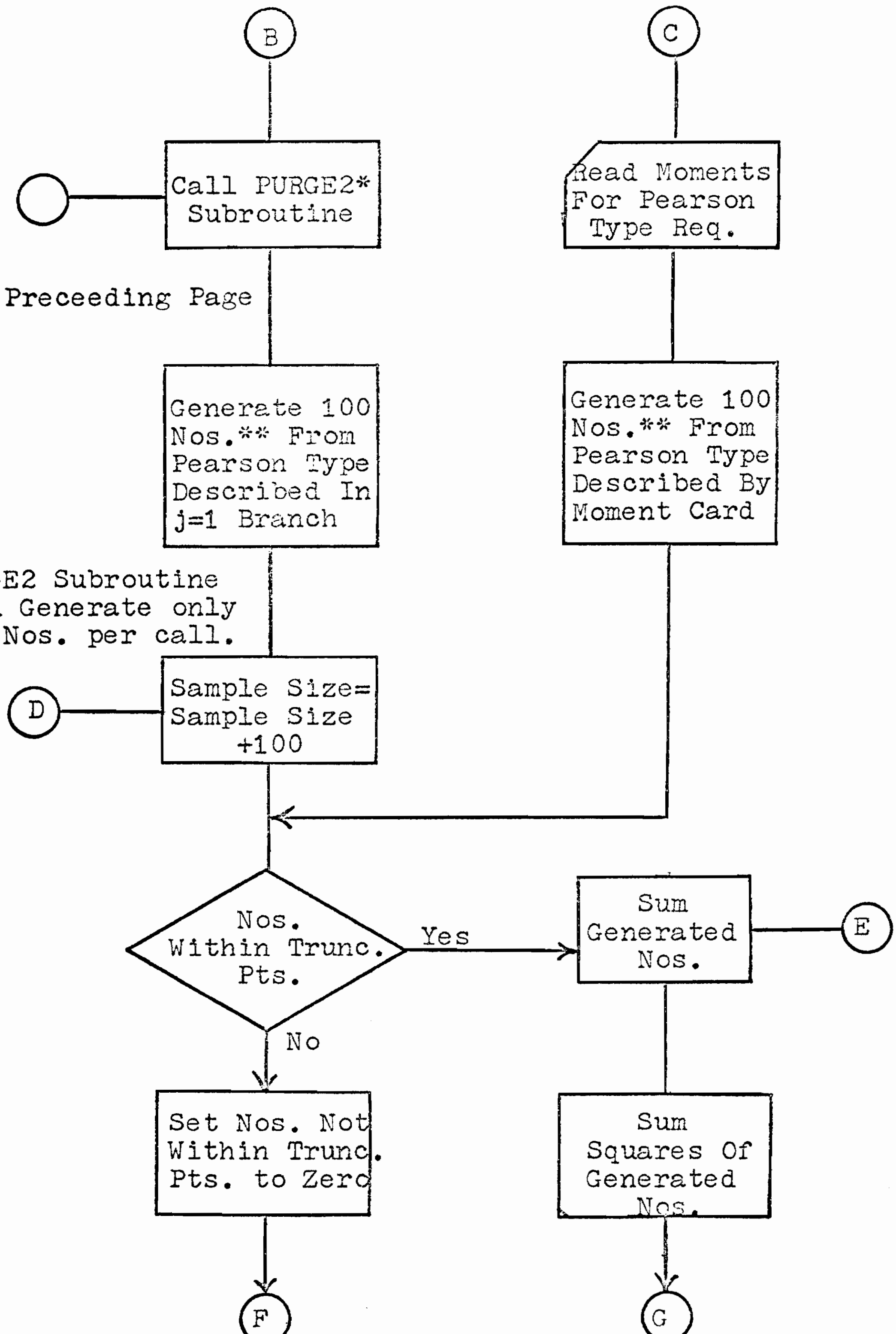
George Carlton Hayles

IX. APPENDIX

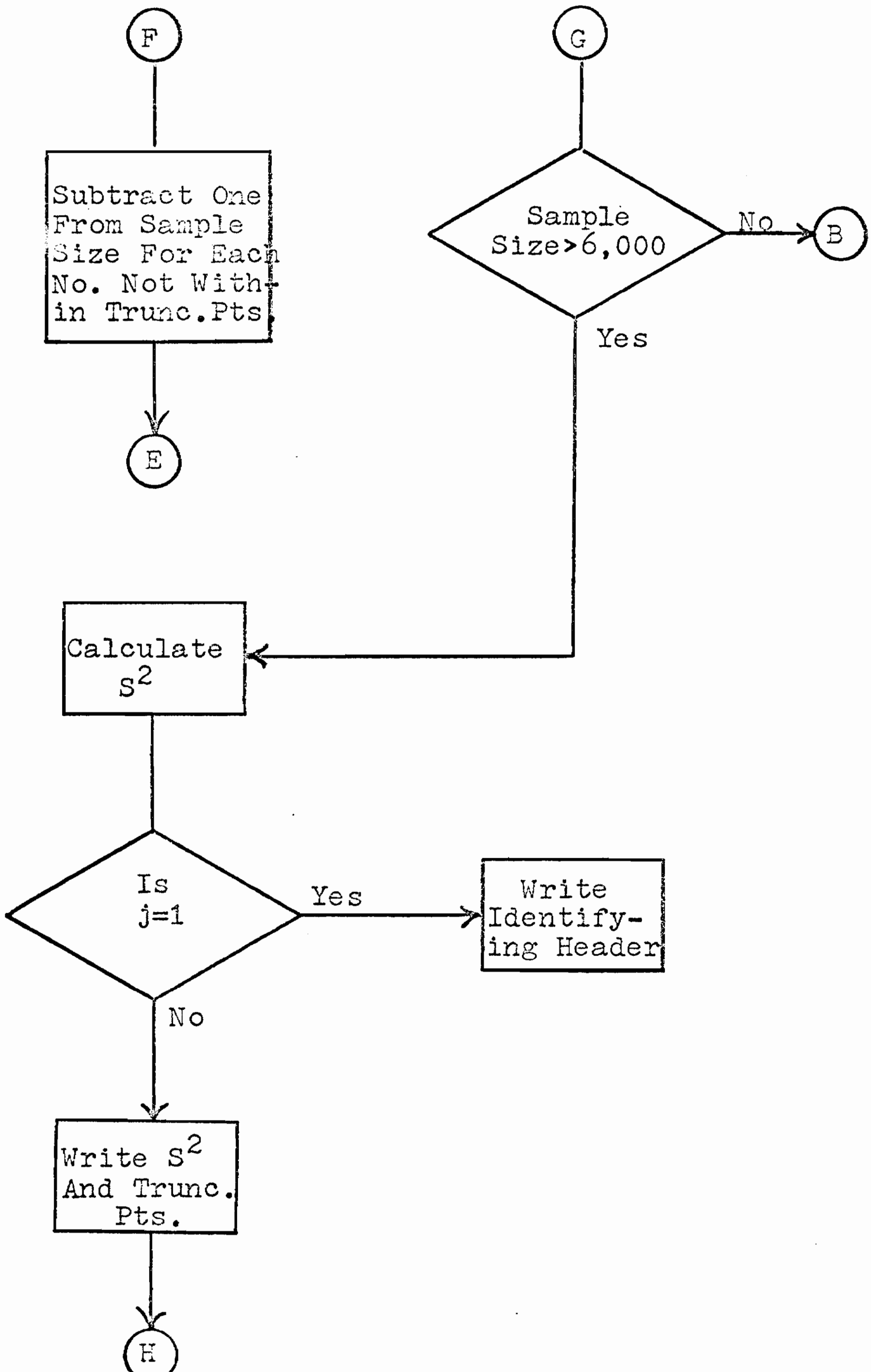


*Complete Flow chart and FORTRAN Source List For PURGE2 Subroutine may be found in Thomas(7).

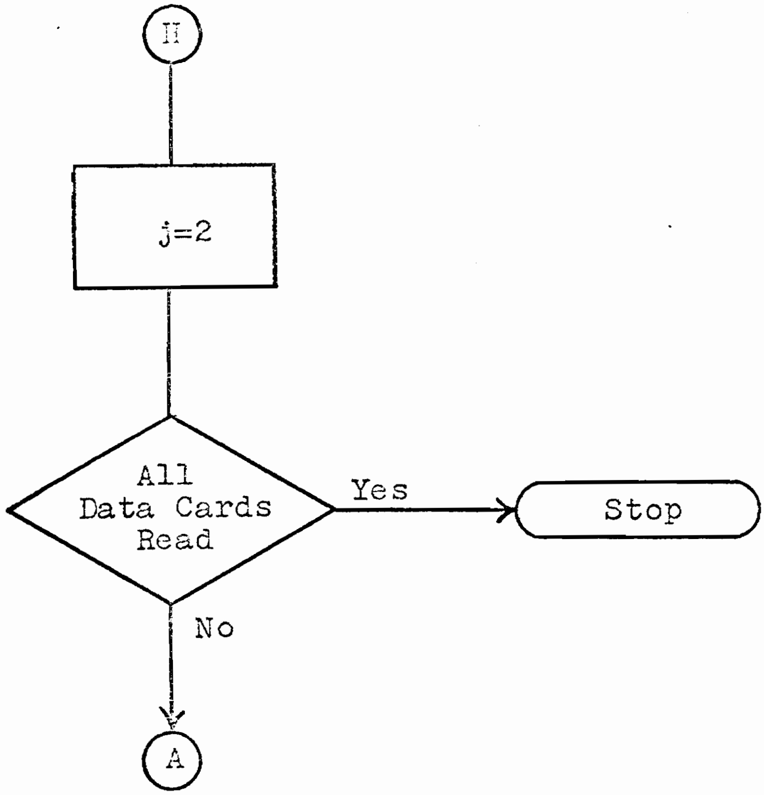
Flow Chart for Pearson Type I U-shaped Distribution



Flow Chart For Pearson Type I U-shaped Distribution, Cont'd



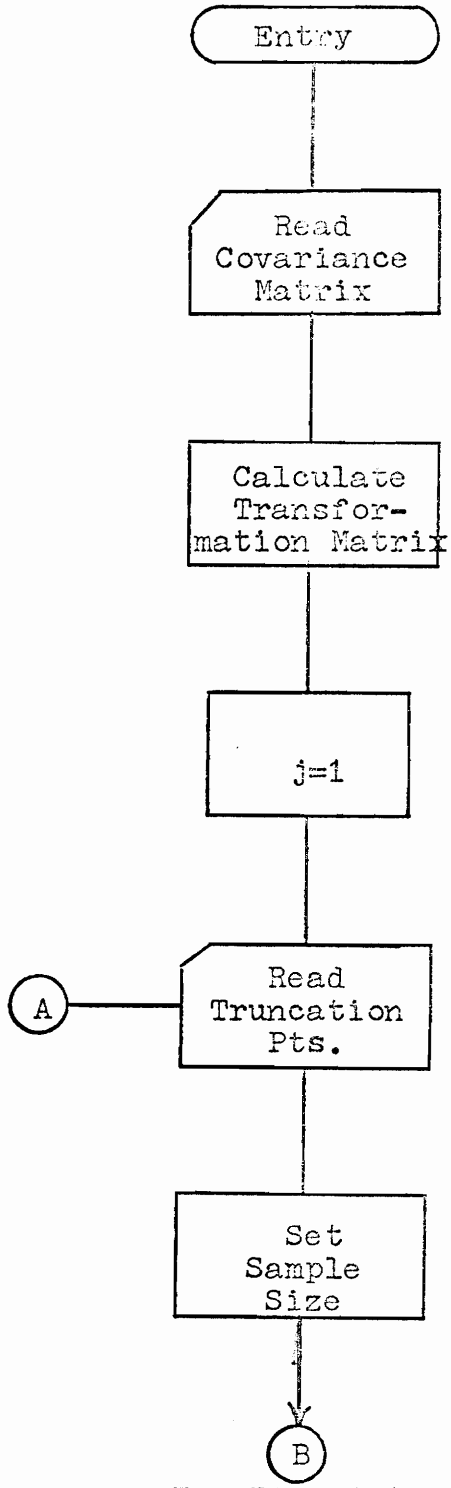
Flow Chart For Pearson Type I U-shaped Distribution, Cont'd



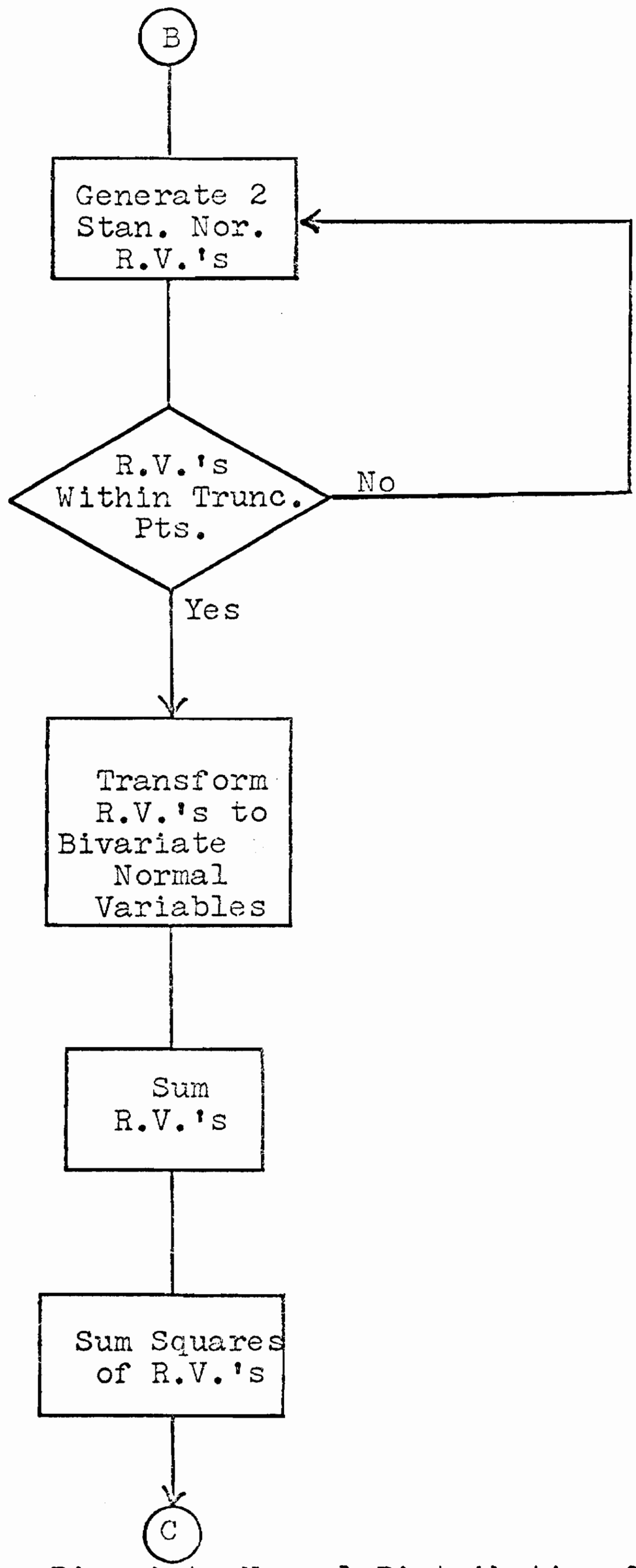
Flow Chart For Pearson Type I U-shaped Distribution, Cont'd

```
      COMMON/VPI001/NUMBR
      READ(5,100)NUMBR
  100 FORMAT(1X,I12)
   10 FORMAT(2F5.2)
      J=1
   99 READ(5,10)B,A
      G=0.
      SX2=0.
      IF(A.GE.B) GO TO 99
      COMMON/Z2/TOM(100)
      IF(J.NE.1) GO TO 2
      CALL PURGE2(1,5)
      GO TO 6
    2 CALL PURGE2(2,5)
    6 NSAMP=0
    7 NSAMP=NSAMP+100
      DO 4 J1=1,100
      IF(TOM(J1).LT.A) GO TO 11
      IF(TOM(J1).GT.B) GO TO 11
      GO TO 3
   11 TOM(J1)=0.
      NSAMP=NSAMP-1
    3 G=G+TOM(J1)
    4 SX2=SX2+TOM(J1)**2
      IF(NSAMP.GT.6000) GO TO 8
      CALL PURGE2(2,5)
      GO TO 7
    8 SV2=SX2-G**2/FLOAT(NSAMP)
      NSAMP1=NSAMP-1
      SV=SV2/FLOAT(NSAMP1)
      IF(J.NE.1) GO TO 9
      WRITE(6,20)
   20 FORMAT(35H      VAR      SAMPLE SIZE      A      B)
    9 WRITE(6,30) SV,NSAMP,A,B
   30 FORMAT(2X,1H0,F7.4,4X,I5,4X,F5.2,4X,F5.2)
      WRITE(6,110)NUMBR
  110 FORMAT(1X,I12)
      END FILE 6
      IF(SV) 97,1,1
    1 J=2
      GO TO 99
   97 STOP
      END
```

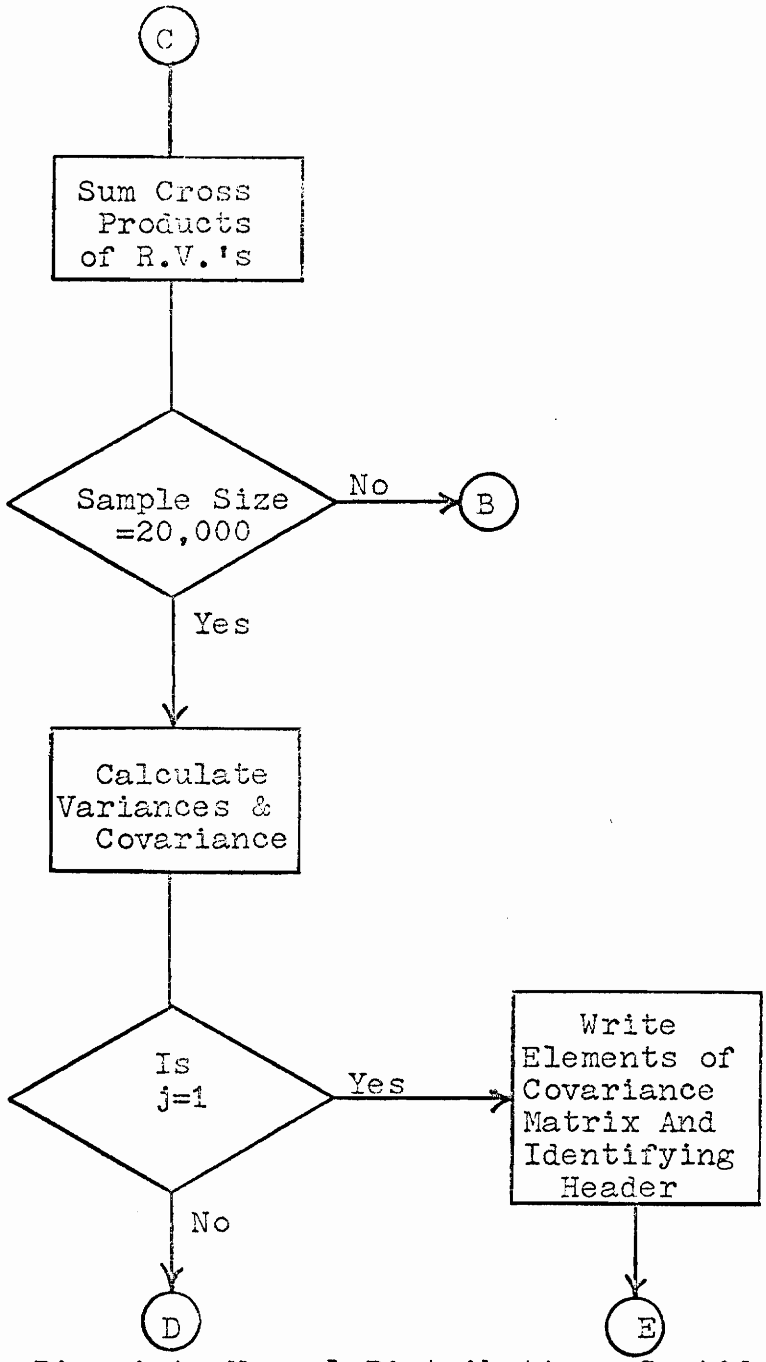
FORTRAN SOURCE LIST FOR PEARSON TYPE I U-SHAPED DISTRIBUTION



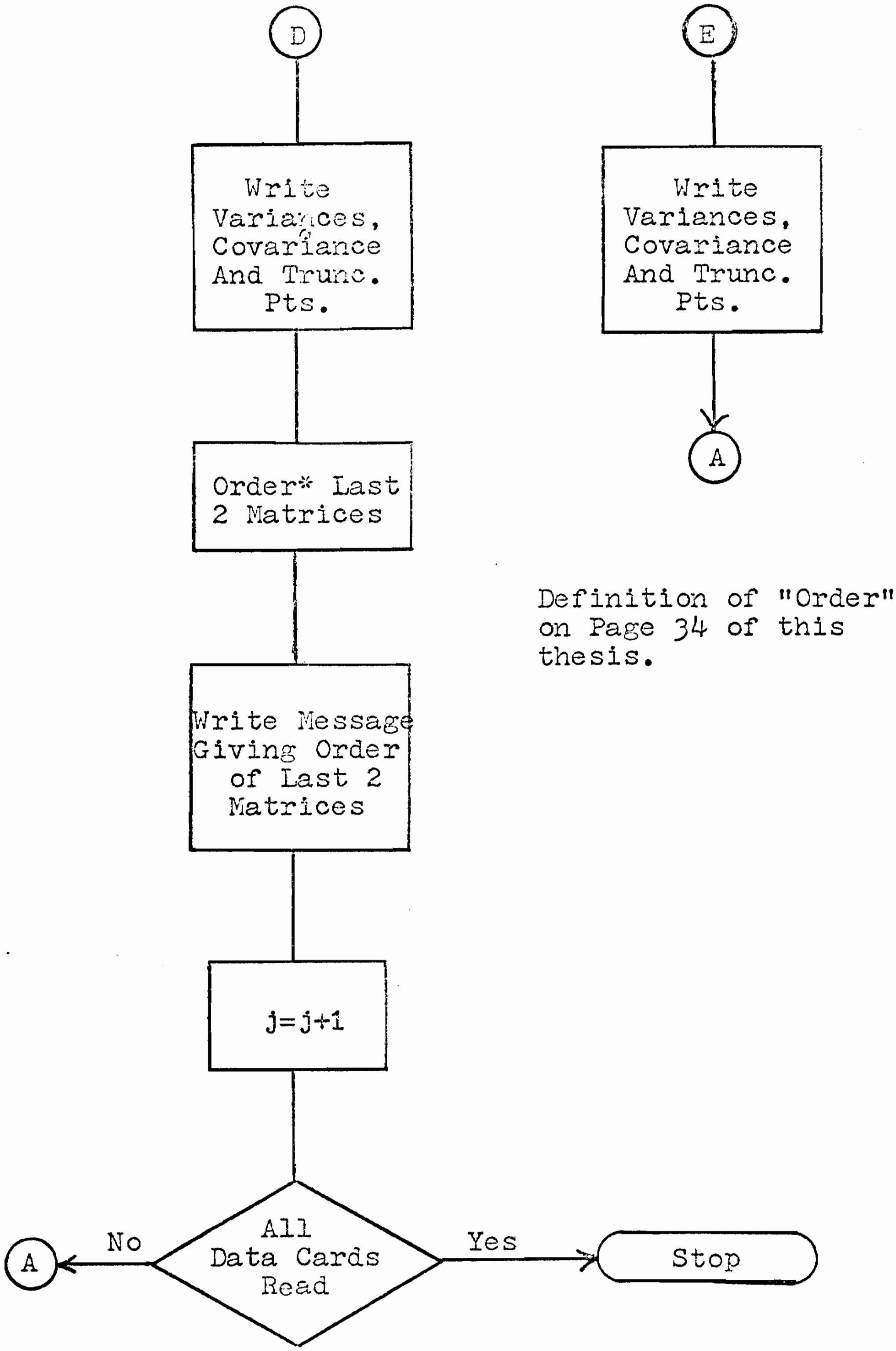
Flow Chart For Bivariate Normal Distribution



Flow Chart For Bivariate Normal Distribution, Cont'd



Flow Chart For Bivariate Normal Distribution, Cont'd



Flow Chart For Bivariate Normal Distribution, Cont'd

```
      DIMENSION S(2),Z(2,1),V11R(100),V12R(100),V22R(100)
      COMMON/VPI002/RANDOM
      READ(5,100)RANDOM
  100 FORMAT(1X,I12)
      READ(5,20)V11,V12,V22
   20 FORMAT(3F7.4)
      V21=V12
      T11=SQRT(V11-V21**2/V22)
      T12=V21/SQRT(V22)
      T21=0.
      T22=SQRT(V22)
      J=1
   11 READ(5,10)A1,A2,B1,B2
      GZ1=0.
      GZ2=0.
      SSZ1=0.
      SSZ2=0.
      GZ1Z2=0.
      NI2=20000
      NI2S1=NI2-1
   10 FORMAT(4F5.2)
      K=0
      L=0
   25 K=K+1
      S(1)=RNOR(X)
      S(2)=RNOR(X)
    4 Z(2,L)=T22*S(2)
      IF((Z(2,L).LT.B1).OR.(Z(2,L).GT.B2)) GO TO 3
      GO TO 7
    3 S(2)=RNOR(X)
      GO TO 4
    7 Z(1,L)=T11*S(1)+T12*S(2)
      IF((Z(1,L).LT.A1).OR.(Z(1,L).GT.A2)) GO TO 6
      GO TO 5
    6 S(1)=RNOR(X)
      GO TO 7
    5 GZ1=GZ1=Z(1,L)
      GZ2=GZ2=Z(2,L)
      GZ1Z2=GZ1Z2+Z(1,L)*Z(2,L)
      SSZ1=SSZ1+Z(1,L)**2
      SSZ2=SSZ2+Z(2,L)**2
      IF(K.NE.NI2) GO TO 25
      V11R(J)=(SSZ1-GZ1**2/FLOAT(NI2))/FLOAT(NI2S1)
      V12R(J)=(GZ1Z2-GZ1*GZ2/FLOAT(NI2))/FLOAT(NI2S1)
      V22R(J)=(SSZ2-GZ2**2/FLOAT(NI2))/FLOAT(NI2S1)
```

FORTRAN SOURCE LIST FOR BIVARIATE NORMAL DISTRIBUTION

```
      IF(J.NE.1) GO TO 9
      WRITE(6,90)V11,V12,V22
   90 FORMAT(21H COVARIANCE MATRIX IS,3F7.4,39H FOR 11,12,
     1AND 22 ELEMENTS RESPECTFULLY)
      WRITE(6,30)
   30 FORMAT(62H    V11      V12      V22      A1    A2    B1
     1    B2)
    9 WRITE(6,40)V11R(J),V12R(J),V22R(J),A1,A2,B1,B2
   40 FORMAT(1H0,3F12.8,1X,F5.2,1X,F5.2,1X,F5.2,1X,F5.2)
      R=V12R(J)/SQRT(V11R(J)*V22R(J))
      R2=R**2
      WRITE(6,80)R2,R
   80 FORMAT(9H RSQUARE=,F12.8,3H R=,F12.8)
      IF(J.EQ.1) GO TO 13
      DIAG1=V11R(J)-V11R(J-1)
      DIAG2=V22R(J)-V22R(J-1)
      DET=DIAG1*DIAG2-(V12R(J)-V12R(J-1))**2
      IF(DIAG1.GT.0..AND.DIAG2.GT.0..AND.DET.GT.0.) GO TO 12
      WRITE(6,50)
   50 FORMAT(120H
     1            DIAG1       DIAG2       DET     VR2-VR1 IS NOT
     1POS.DEF.)
      GO TO 15
   12 WRITE(6,60)
   60 FORMAT(116H
     1            DIAG1       DIAG2       DET     VR2-VR1 IS
     1POS.DEF.)
   15 WRITE(6,70) DIAG1,DIAG2,DET
   70 FORMAT(1H ,63X,3F12.8)
      WRITE(6,110)RANDOM
  110 FORMAT(1X,I12)
      END FILE 6
   13 J=J+1
      GO TO 11
      END
```

FORTRAN SOURCE LIST FOR BIVARIATE NORMAL DISTRIBUTION, CONT'D

VARIANCES OF SOME TRUNCATED DISTRIBUTIONS

FOR VARIOUS POINTS OF TRUNCATION

by

George Carlton Hayles

ABSTRACT

The purpose of this study is to examine variances in the case of distributions obtained by truncating a given distribution at various points. In particular, the truncated distributions are restricted to nested increasing intervals, and the question is posed whether the variances of these distributions are monotonically increasing. The answer to this question is relevant to the use of conditional information for purposes of estimation and prediction.

Several tables are presented in the thesis which provide evidence of the property of monotonic variance for nested increasing intervals of truncation in the case of univariate distributions. The Monte Carlo procedure is used to determine a table of standard deviations for the standard normal distribution with the same points of truncation reported by Clark(2). Clark's table is given intact, and it is used in comparison with the new table reported here as a check on the Monte Carlo procedure used in the present study.

Distributions other than the standard normal distribution are examined as well, namely, a Pearson U-shaped distribution and a bimodal distribution consisting of a mixture of two Pearson distributions. Graphs of the U-shaped and bimodal distributions are given.

A section is given in which dispersion for a bivariate case is examined in terms of the bivariate normal distribution. An interesting trend among the covariance matrices is observed in the data reported in that section.

A separate computer program for each type of distribution was written and used to calculate the variances of the truncated distributions. FORTRAN programs and flow charts are presented in the Appendix. Explanation of the tables and procedures used to calculate the entries in the body of each table are given in each section as well as some discussion of the results presented.