DOCUMENT RESUME

ED 082 492 EM 011 461

AUTHOR	Beaujon, Hendrik Johannes
TITLE	An Interactive, Graphical Display System for Illustrating Elementary Properties of Statistical Distributions.
INSTITUTION	North Carolina Univ., Chapel Hill. Dept. of Computer Science.
PUB DATE	70
NOTE	61p.; Thesis submitted to the Department of Computer and Information Science of the University of North Carolina
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Computer Assisted Instruction; Computer Based Laboratories; *Computer Graphics; Higher Education; Input Output Devices; *Interaction; Man Machine Systems; Masters Theses; *Probability; *Statistics; Student Attitudes
IDENTIFIERS	CAI; IBM 2250 Display Unit Model 1; IBM Systems 360 Model 40 Computer

ABSTRACT

A computer-assisted instruction (CAI) project designed a system to teach the elementary properties of statistical distributions in a probability course. It was used to reinforce teacher material with graphic displays and to provide laboratory exercises, the objectives being to promote easier grasp and retention of subject matter. The system developed was capable of displaying: 1) probability functions and cumulative distribution functions for discrete distributions; 2) probability density functions and cumulative distribution functions for continuous distributions; and 3) random samples, with resulting frequency distributions. The hardware utilized included an IBM 2250 Display Unit Model 1 attached to an IBM Systems/360 Model 40 computer with 256K bytes of memory. Trial use indicated that the system aided teaching; students reacted favorably, and they were able to complete the laboratory exercises. The system was easy to use and provided fast and reliable responses. Also, the interactive graphics display mode possessed additional advantages over drawings of distributions. These included: 1) the variation of parameters, thus allowing a huge number of displays; 2) superposition, to compare shapes of distributions; 3) zooming capability to allow variation of an abscissa or ordinate; 4) increased student motivation. (PB)

University of North Carolina at Chapel Hill

Department of Computer and
Information Science

AN INTERACTIVE, GRAPHICAL DISPLAY SYSTEM FOR ILLUSTRATING ELEMENTARY PROPERTIES OF STATISTICAL DISTRIBUTIONS

by

Hendrik Johannes Beaujon

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements of the degree of Master of Science in the Department of Computer and Information Science

Chapel Hill

1970

Approved by:

Advisor ______________________

HENDRIK JOHANNES BEAUJON. An Interactive, Graphical Display System for Illustrating Elementary Properties of Statistical Distributions. (Under the direction of DR. PETER CALINGAERT)

ABSTRACT

The purpose of this computer project was to design a system to display statistical distributions. The hardware utilized was an IBM 2250 Display Unit Model 1 attached to an IBM Systems/360 Model 40 computer with 256K bytes of memory.

The system was designed as a teaching tool. Primary desiderata were ease of use, speed of response, and reliability.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to Dr. P. Calingaert and Dr. D. Guthrie for their guidance in the preparation of this thesis, and to the International Business Machines Corporation which made available, under a research contract between that corporation and the University of North Carolina at Chapel Hill, the IBM 2250 Display unit used.

THE PURPOSE OF COMPUTING IS INSIGHT, NOT NUMBERS.

R. W. Hamming

CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

1.1. Purpose of the System

This interactive, graphical display system is intended as an aid in teaching the elementary properties of statistical distributions to students in an introductory probability course. The author visualizes two essentially different ways in which the system could usefully be incorporated in a college level course:

(1) In a classroom situation where the teacher reinforces his material with graphical displays. In this situation the teacher is in command of the system.

(2) In a laboratory situation where one or more students, possibly supervised, are in command of the system, and go through a number of prepared exercises.

In either situation the sequence of graphical displays will have to be carefully prepared in advance and, in the case of the classroom situation, inserted at strategic points during the class. The system derives its usefulness for a great part from its ability to answer exploratory questions. Thus these should be highly encouraged and, whenever possible, answered by means of graphical displays.

1.2. Objectives

The system has the following graphical capabilities.

(1) For a number of standard discrete statistical distributions one can display the probability function or the cumulative distribution function.

(2) For a number of standard continuous statistical distributions one can display the probability density function or the cumulative distribution function.

(3) From a number of standard statistical distributions a random sample can be drawn and the resulting frequency distribution displayed.

(4) Any of the foregoing displays can be exhibited simultaneously.

Under (1), (2), and (3) one can freely vary the parameters of the distributions. Under (3) one can vary the sample size, the number of samples and the width of the interval for the resulting frequency distribution. The system is particularly well suited to illustrate limit properties of distributions, e.g., that the Poisson distribution tends to the normal distribution as its mean gets large. The sampling capability was incorporated in order to illustrate the Central Limit Theorem. The resulting frequency distribution is really a distribution of the normalized mean of each sample. The histogram should thus rougly resemble a normal distribution.

The intention is that the introduction of graphical displays will result in easier grasp and longer retention of the subject matter.

II. SYSTEM ENVIRONMENT

2.1. Hardware

The system was programmed for the IBM 2250 Display Unit Model 1. This unit is attached to an IBM System/360 Model 40 computer (with 256K bytes of memory) via a selector channel. Figure 1 shows the current hardware configuration at the University of North Carolina at Chapel Hill.

The main feature of the 2250 is a cathode-ray tube (CRT) on which computer programmed graphic and alphanumeric information is displayed. The area on the face of the CRT on which the images may be displayed is a twelve inch by twelve inch grid. This area is a square consisting of discrete points. These equally spaced points, 1024 in both the vertical and the horizontal direction, are 1024 x 1024 in number.

The 2250 used for this system is a basic 2250 unit plus the following features:

(1) Absolute vectors and control. This feature allows vectors of any length to be drawn at any angle on the CRT. Without this feature the basic 2250 can display only horizontal and vertical vectors of unrestricted length, and 45 degree vectors of

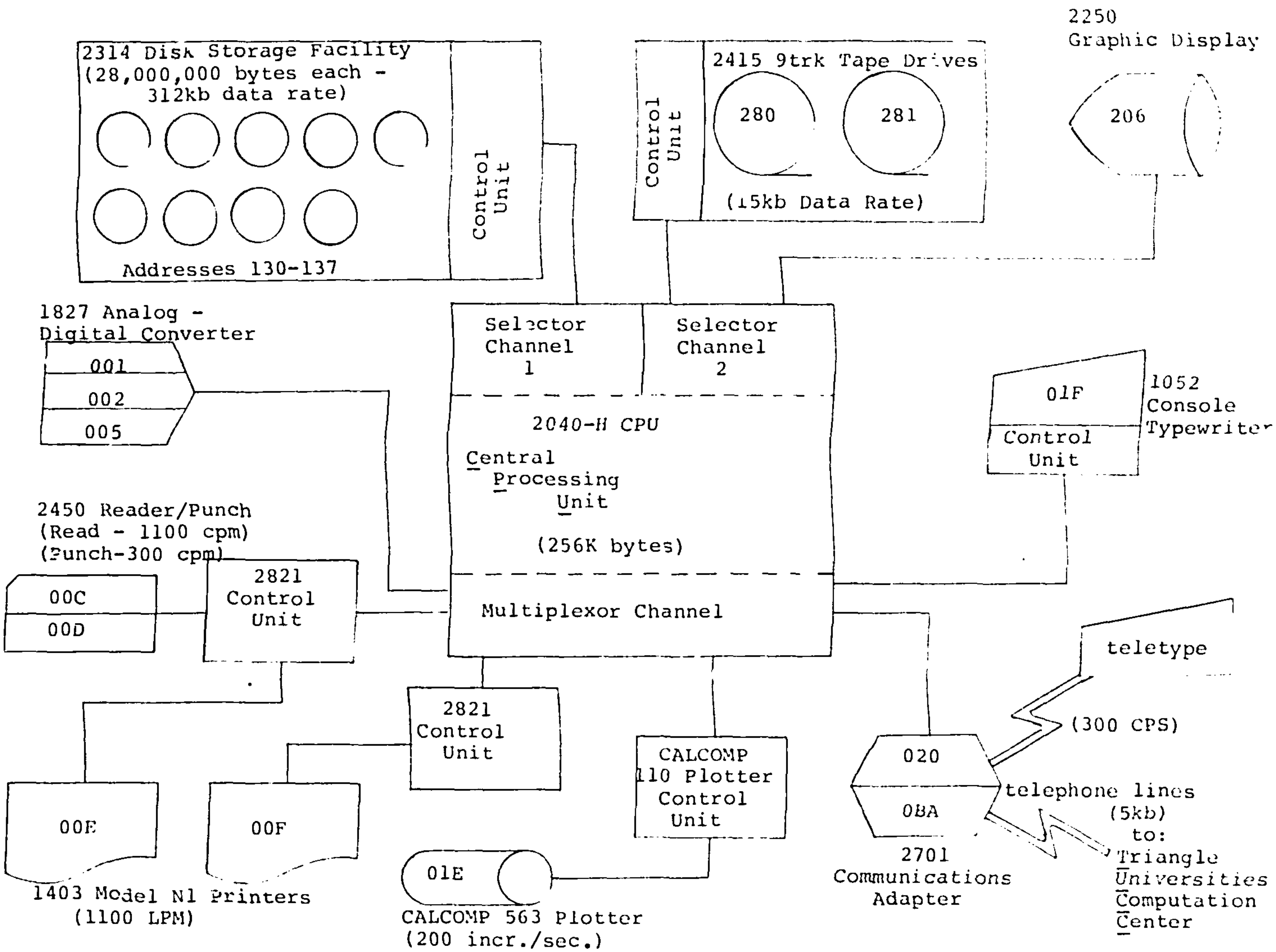

Figure 1. Computation Center - University of North Carolina at Chapel Hill IBM System/360 Hardware Configuration November 1969

limited length.

(2) 2250 local buffer. Use of this 8K buffer to regenerate displays allows the 2250 to operate concurrently with the computer.

(3) Alphanumeric keyboard. This is a typewriter-like keyboard which can be used to enter messages into the 2250 buffer and onto the CRT display area.

(4) Programmed function keyboard. This consists of thirty-two pushbutton keys each with its own indicator light. Each key can be made to correspond to a set of instructions in the computer program. The keyboard has a replaceable overlay which labels each key and which is itself program identifiable by an eight-bit code notched into one edge.

(5) Character generator. This allows the 2250 to translate a System/360 character representation and trace the character on the CRT display area. A set of 63 characters, each available in two sizes, is used.

(6) Lightpen detect. This allows the user to enter information into the system by pointing a pen-like device at a lighted portion of the CRT and depressing the associated foot switch. Control is transferred to the computer program.

A complete description of the IBM 2250 Model 1 may be found in Reference 4.

2.2. Software

The statistical display system consists of 26 load modules totaling approximately 30K bytes of memory. The advantages of such a modular structure are well known. It is particularly suited to this type of system containing many self-contained units, e.g., the subroutines to calculate the function values of the statistical distributions. The reader is referred to Chapter IV for a more detailed discussion of the logical structure of the computer system. Chapter IV also contains descriptions of the individual subprograms.

The multiprogramming environment at the Computation Center of the University of North Carolina allows the display system to be used concurrently with batch processing without appreciably affecting the response time at the 2250 unit., i.e., the waiting time from the moment a particular display is requested to the instance of appearance on the screen. This time is, of course, also a function of the amount of computation involved.

The system is invoked by the following Job Control Statements in the OS (release 16) job stream.

```
//JOBNAME JOB (UNC.CS.S6246,,T),'BEAUJON,J',MSGLEVEL=1,    *
//             PRTY=10,CLASS=G,TYPRUN=HOLD
//JOBLIB   DD  UNIT=DISK,DISP=SHR,DSNAME=UNC.IS.F7027.     *
//             ISDEPT.GRAPHICS
//         EXEC PGM=UCSTAT
//DISPLAY DD   UNIT=DISPLAY
```

```
//SYSUDUMP  DD SYSOUT=A
```

Typically, the user will request the computer operator to release 'BEAUJON' which is a catalogued procedure containing the preceding set of Job Control Statements.

The statistical display system can be implemented in any environment supporting OS and a 2250 Unit Model 1. The least important feature of the 2250 for this particular application is the lightpen. It would also be possible to do without the 2250 local buffer. The modifications required would be minor in case of absence of the lightpen. Absence of the 2250 local buffer would necessitate more extensive changes. OS is a *sine qua non*.

III. USER INSTRUCTIONS

3.1. Introduction

The user enters information into the system in one of the following ways:

(1) by pressing one of the keys of the programmed function keyboard;

(2) by typing in information via the alphanumeric keyboard;

(3) by pointing the lightpen at a section of the screen and pressing the associated foot switch.

The system may respond by:

(1) providing the desired display;

(2) removing the appropriate display;

(3) entering into the desired mode;

(4) delivering an error message;

(5) terminating.

Typically, the user will go through an initialization procedure in order to obtain a grid so as to define the extent of his abcissa and ordinate. After this the user enters the parameters of the desired distribution and presses the programmed function key (pfk) pertaining to that distribution. Depending on the current mode of the system the resulting plot is a probability distribution function (density function for continuous distributions), a cumulative distribution function, or a frequency distribution depicting results of a sample. The system may be put in a particular mode by pressing the appropriate pfk. Displays may be removed via the programmed function keyboard or by means of the lightpen.

3.2. Generating Displays

3.2.1. Detailed Procedure Description

Following is a step by step description of the procedure necessary to display distributions.

(1) Starting point. The graphics package has been released. The general instructions are displayed on the screen and pfk INST is lighted.

(2) Press pfk RESET: the residing grid parameters appear. The display should look like Figure 2a. (If this is not the case something is wrong and the operator should be notified over the intercom.) In the following discussion, let

xlo be the lowest X value to be displayed;
xhi be the highest X value to be displayed;
xinc be the interval at which vertical grid
  lines will be drawn;
ylo be the lowest Y value to be displayed;
yhi be the highest Y value to be displayed;

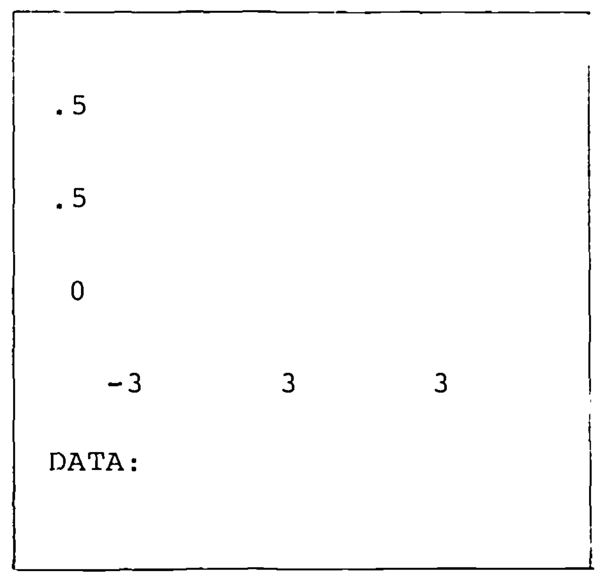

Figure 2a. 2250 Screen following RESET.

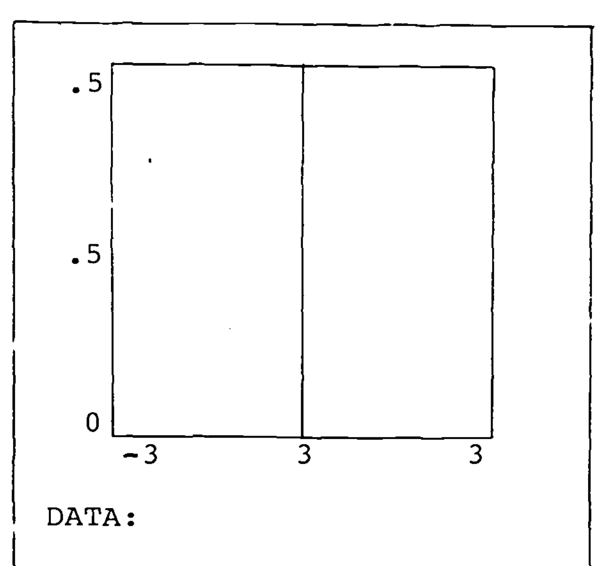

Figure 2b. 2250 Screen following RESET,GRID.

yinc be the interval at which horizontal grid lines will be drawn.

Thus RESET assigns the following values to the grid parameters: xlo = -3; xhi = 3; xinc = 3; ylo = 0; yhi = .5; yinc = .5 .

(3) If the residing grid parameters are satisfactory, i.e., if the distribution to be drawn will fall within these limits: press pfk GRID. The grid appears and should look like Figure 2b. If the residing grid parameters are not satisfactory these (and therefore the ensuing grid) may be altered by:

(4) Entering new grid parameters. Pressing the JUMP key on the alphanumeric keyboard has the effect of moving the cursor (the bright underscore on the screen) according to the pattern shown in Figure 3. Grid parameters may be changed by simply typing in new numbers while the cursor is at the corresponding parameter location, e.g., if xinc is to be changed to 6 and the cursor is located somewhere to the right of the word DATA, pressing the JUMP key three times will place the cursor at the desired location. One then types a 6 on the alphanumeric keyboard. After the grid parameters have been altered the grid is obtained by pressing pfk GRID. Note that it is possible to press GRID at any time. Distributions currently

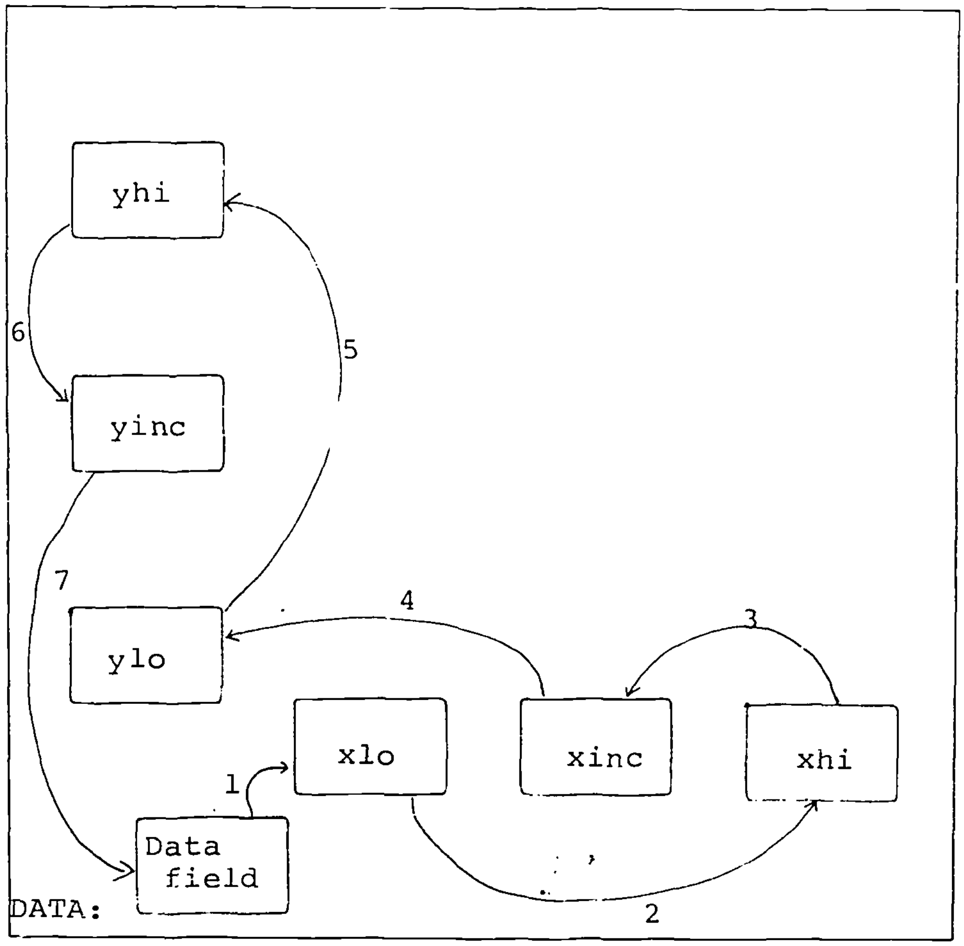

Figure 3. Jump Pattern of the Cursor

displayed, if any, are erased and a new grid corresponding to the new grid parameters is drawn.

(5) Mode. Before displaying a distribution the user should ascertain that the system is in the mode corresponding to the desired display, i.e., if a probability density is desired the system should be in the density mode. One of the pfk's PDF, CDF or SAMPLE is always lit, indicating the current mode. PDF corresponds to the probability density mode, CDF corresponds to the cumulative distribution mode and SAMPLE to the sample mode. The system stays in a particular mode until one of the other mode pfk's is pressed. The exception to this is that RESET puts the system back into the density mode. Thus the user need only press the pfk PDF or CDF to have the system enter that mode. For the sample mode additional information has to be supplied. Before pressing SAMPLE three parameters have to be typed in at the data field:

(i) the size of each sample;

(ii) the number of samples;

(iii) the interval width for the frequency distribution.

Table 1 is a summary of the displays produced when the system is in the various modes.

Table 1. Computations of the Various Modes

	discrete distributions	continuous distributions
PDF	$P[X = x]$	$p(x)$
CDF	$\sum_{x}^{a} P[X = x]$	$\int_{-\infty}^{a} p(x)\,dx$
SAMPLE	frequency distribution of $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	

where $\bar{x}$ = the sample mean;

μ = the mean of the distribution sampled from;

σ = the standard deviation of the distribution sampled from;

n = the sample size.

(6) The user is now ready to define the particular distribution. This is done by

(i) Typing in the parameters of that distribution. See Table 2 for the number and meaning of the parameters corresponding to a particular distribution.

(ii) Pressing the pfk bearing the label of the distribution to be plotted.

To superpose distributions within the same mode and on the same grid, step (6) need only be repeated. To superpose distributions on the same grid but in a different mode, steps (5) and (6) have to be repeated. At a given time the display may consist of a superposition of all types of distributions. Grids can not be superposed. A maximum of seven distributions may be displayed at any one time. An attempt to display more than seven results in an error message being displayed.

Entering grid parameters: The cursor should be located at the grid parameter to be changed. One can accomplish this by pressing the JUMP key on the alphanumeric keyboard. The parameter is entered as a decimal number, possibly preceded by a minus sign. xinc or yinc should not be negative. Grid parameters are not followed by a comma.

Table 2. Parameters to be entered for a Distribution

(see distribution definitions)

pfk	Distribution	Parameter(s)		
12	discrete uniform	a	b	n
13	binomial	n	p	
14	Poisson	λ		
15	hypergeometric	N	n	k
16	continuous uniform	a	b	
17	triangular	a	b	
18	normal	μ	σ^2	
19	gamma	α	β	
20	exponential	θ		
21	Cauchy	a		
22	beta	α	β	
23	Student's t	n		
24	chi square	n		
25	F	m	n	

Entering sample parameters: The cursor should be located at the DATA field. The parameters to be entered, in order, are: sample size, number of samples, and width of interval.

Example: The samples to be drawn are of size 5. Furthermore, 50 of these samples (each of size 5) should be drawn. The resulting frequency distribution should have an interval width of .5.

Type in at DATA field: 5,50,.5,

Note that in order to draw a sample from a distribution the parameters of that distribution have to be entered and also the pfk pertaining to that distribution must be pressed. SAMPLE does not draw a sample but merely puts the system in that mode. Sampling thus consists of the following four steps:

(i) entering sample parameters;

(ii) pressing SAMPLE;

(iii) entering distribution parameters;

(iv) pressing the distribution pfk.

Entering distribution parameters: The cursor should be located at the DATA field. The number and meaning of the parameters depend on the distribution. See Table 2.

Example: a) for Normal (0,1) type 0,1,

b) for Poisson (7) type 7,

Note that the distribution parameters and the sample parameters are each followed by a comma, including the last.

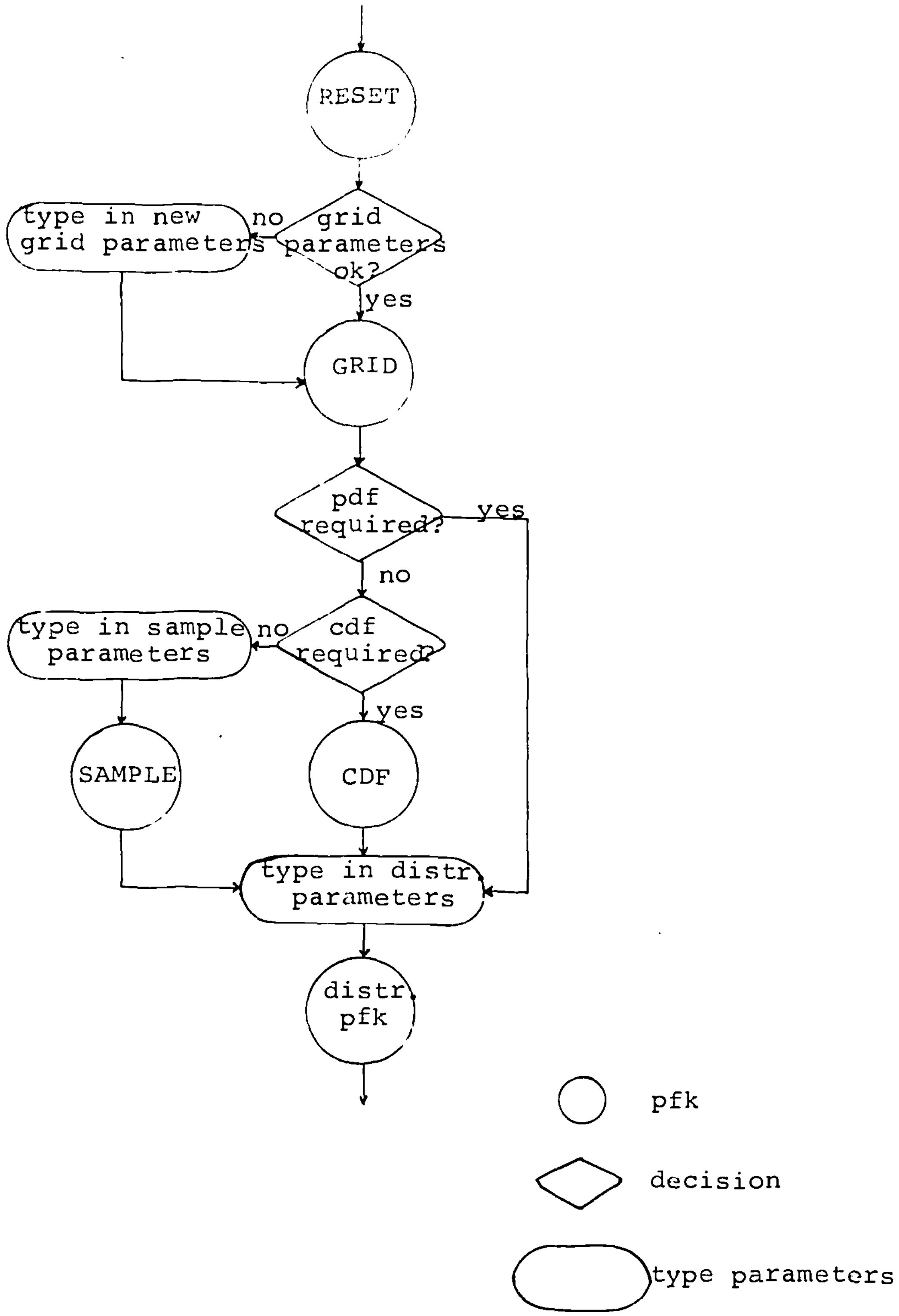

Figure 4. Schematic Decision Flow for Obtaining Displays

Erasing distributions

Distributions may be erased as follows:

(1) By pressing RESET. All distributions are removed, and the default grid parameters appear. Before proceeding to new distributions GRID must be pressed.

(2) By pressing GRID. All distributions are removed, and the grid corresponding to the current grid parameters appears.

(3) By pressing REML: this erases the distribution last plotted.

(4) By pressing REM NL; this erases the distribution plotted next to last.

(5) By pointing the lightpen at any portion of a particular distribution and stepping on the foot switch. This erases that distribution.

Summarizing: (1) and (2) remove all distributions, while (3), (4), and (5) remove one particular distribution.

3.2.2. Action of the Programmed Function Keys

Following is a detailed description of the action of the programmed function keys. In each case the key last pressed is lit. Also one of the keys PDF, CDF or SAMPLE is always lit indicating in which mode the system is. The removable overlay is color coded to indicate functional similarity among the keys.

Number	Label	Description
00	INST	The general instructions are displayed. This may be done at any time. The display visible just prior to the instructions may be redisplayed by pressing the REGEN pfk.
01	RESET	The system returns to the permanent display with its residing grid parameters. This may be done at any time, including after INST has been pressed. It is thus not mandatory in that case to follow INST by REGEN. Following RESET only four pfk's are active: (1) INST; (2) RESET; (3) GRID; (4) END. Other pfk's are ignored. RESET always puts the system into the density mode.
02	GRID	The grid is displayed according to the current grid parameters. Immediately following RESET no grid parameters need be entered if the residing grid parameters are satisfactory. After changing the grid parameters, either by using pfk's 06 through 09 or by using the alphanumeric keyboard, GRID should be pressed. The distributions currently displayed disappear but the information is retained. The distributions may be redisplayed on the new scales

Number	Label	Description
		by pressing REDRAW. The information concerning the distributions is destroyed if REDRAW is not pressed following GRID.
03	REDRAW	The distributions visible just prior to GRID being pressed are redrawn on the new scales. This may be done at any time. It is, of course, a waste to do so if the grid parameters were not changed.
04	PDF	The system enters into the density mode. See section 3.2.1. for the exact differences among the three modes. The system stays in this mode until either CDF or SAMPLE is pressed. This may be done at any time so that any current display may contain a mixture of densities, cumulative distribution functions and sample results.
05	REGEN	This regenerates the display visible just prior to pressing INST. REGEN is ignored if INST was not the preceding pfk.
06	X/2	The X grid parameters are divided by 2. GRID must then be pressed. After this one might wish to use REDRAW.
07	Y/2	The Y grid parameters are divided by 2. See X/2.
08	X.2	The X grid parameters are multiplied by 2. See X/2.

Number	Label	Description
09	Y.2	The Y grid parameters are multiplied by 2. See X/2. A combination of pfk's 06 through 09 may be pressed before pressing GRID.
10	CDF	The system enters the cumulative distribution mode. Pressing the statistical distribution keys will result in cdf's being drawn.
11	SAMPLE	The system enters the sample mode. In this mode random samples may be drawn from the statistical distributions. The parameters read in when SAMPLE is pressed are respectively: the sample size, the number of samples, and the interval width to be used for the resulting histogram. These parameters are absolutely necessary. They must be entered prior to pressing SAMPLE. See section 3.2.1. for the exact meaning of these parameters.
12	UNIF	Plots the discrete uniform distribution.
13	BINOM	Plots the binomial distribution.
14	POISS	Plots the Poisson distribution.
15	HYPER	Plots the hypergeometric distribution.
16	UNIF	Plots the continuous uniform distribution.
17	TRIAN	Plots the triangular distribution.
18	NORMAL	Plots the normal distribution.
19	GAMMA	Plots the gamma distribution.
20	EXPO	Plots the exponential distribution.
21	CAUCHY	Plots the Cauchy distribution.

Number	Label	Description
22	BETA	Plots the beta distribution.
23	STUD T	Plots Student's t distribution.
24	CHI 2	Plots the chi square distribution.
25	F	Plots the F distribution.
26		Not implemented.
27		Not implemented.
28		Not implemented.
29	REM NL	Given that there are n distributions plotted, REM NL removes the next to last plotted. If n is less than two REM NL is ignored.
30	REML	Given that there are n distributions plotted, REML removes the last. REML is ignored if n is zero. REML and REM NL may be used repeatedly and intermixed fashion.
31	END	This terminates the system. The program comes to a halt and the 2250 screen goes blank. If this pfk is pressed accidentally the user must request the computer operator to release 'BEAUJON' again.

3.2.3. Programmed Distributions

Following are the definitions used of the statistical distributions presently incorporated in the system. The order of the distribution definitions corresponds to their corresponding pfk's. The first four are discrete distributions.

The rest are continuous. Table 2 lists the parameters to be entered for each distribution. The order in which the distribution parameters are entered on the alphanumeric keyboard is extremely important, of course, since the system has no other way of determining the correspondence between numbers entered and the parameters.

Distribution	Definition	
discrete uniform	$P[x=x] = \frac{1}{n}$	n positive integer
		$b > a$
	$x = a + \left(\frac{b-a}{n-1}\right) i$	
	$i = 0,1,2,\ldots,(n-1)$	
binomial	$P[X=x] = \binom{n}{x} p^x (1-p)^{n-x}$	n positive integer
	$x = 0,1,2,\ldots,n$	$0 \leq p \leq 1$
Poisson	$P[X=x] = \frac{e^{-\lambda}\lambda^x}{x!}$	$\lambda > 0$
	$x = 0,1,2,\ldots$	
hypergeometric	$P[X=x] = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$	N,n,k positive integers
		$n \leq N$
	$x = 0,1,2,\ldots,\min(n,k)$	$k \leq N$

Distribution	Definition	
continuous uniform	$f(x) = \frac{1}{b-a}$	$a \leq x \leq b$
	$= 0$	otherwise
triangular	$f(x) = \frac{4(x-a)}{(b-a)}$	$a \leq x < \frac{a+b}{2}$
	$= \frac{-4(x-b)}{(b-a)}$	$\frac{a+b}{2} \leq x < b$
	$= 0$	otherwise
normal	$f(x) = \frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}$	$-\infty < x < \infty$ $\sigma > 0$
Cauchy	$f(x) = \frac{a}{\pi(a^2+x^2)}$	$-\infty < x < \infty$ $a > 0$
gamma	$f(x) = \frac{x^{\alpha}e^{-x/\beta}}{\Gamma(\alpha+1)\beta^{\alpha+1}}$	$0 < x < \infty$ $\alpha > -1$ $\beta > 0$
	$= 0$	otherwise
exponential	$f(x) = \frac{1}{\theta} e^{-x/\theta}$	$0 \leq x < \infty$ $\theta > 0$
	$=0$	otherwise

Distribution	Definition	
chi square	$f(x) = \dfrac{x^{\frac{n-2}{2}} e^{-x/2}}{2^{n/2}\,\Gamma(n/2)}$	$0 \le x < \infty$ $n > 0$
	$= 0$	otherwise
Student's t	$f(x) = \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left(1+\frac{x^2}{n}\right)^{1/2(n+1)}}$	$-\infty < x < \infty$ $n > 0$
beta	$f(x) = \dfrac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}\, x^{\alpha}(1-x)^{\beta}$	$0 \le x \le 1$ $\alpha > -1$ $\beta > -1$
	$= 0$ otherwise	
F	$f(x) = \dfrac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{2}\right)^{\frac{m}{2}} x^{\frac{m-2}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}}$	$m > 0$ $n > 0$ $0 \le x < \infty$
	$= 0$ otherwise	

3.3. Sample Exercises

The exercises contained in this section were designed for use by someone being exposed to the system for the first time. They are intended to aid the familiarization process as well as to illustrate some basic statistical facts.

Figures 5 a) through 5 d) are actual photographs of the 2250 screen after the last step of exercise 1 through Exercise 4 respectively. The set was tried with good results on a small group of students. For a further discussion of this see section 5.2.

The exercises were written and commented so as to be performed by students in a laboratory situation. They could, of course, also be used as a guideline for a teacher planning to use the display system in a classroom demonstration.

Exercise 1 is intended to show how the probability density function of the normal distribution is affected by a change in the mean and/or variance of the distribution. The first density displayed is that of the unit normal which serves as a basis of comparison for densities subsequently displayed.

Exercise 2 is basically the same as Exercise 1 except that it concerns cumulative distribution functions rather than densities.

Exercise 3 shows the relationship between the binomial distribution and the normal distribution. A series of binomial distribution functions is displayed, the last being binomial (16,.5). The instructions then draw the student's attention to the resemblance of this distribution to a normal distribution. He is asked to try and match, by trial and error, a normal distribution to this last binomial distribution. (It is expected that he has previously performed Exercises 1 and 2.) The student is also asked to check the result of his match by

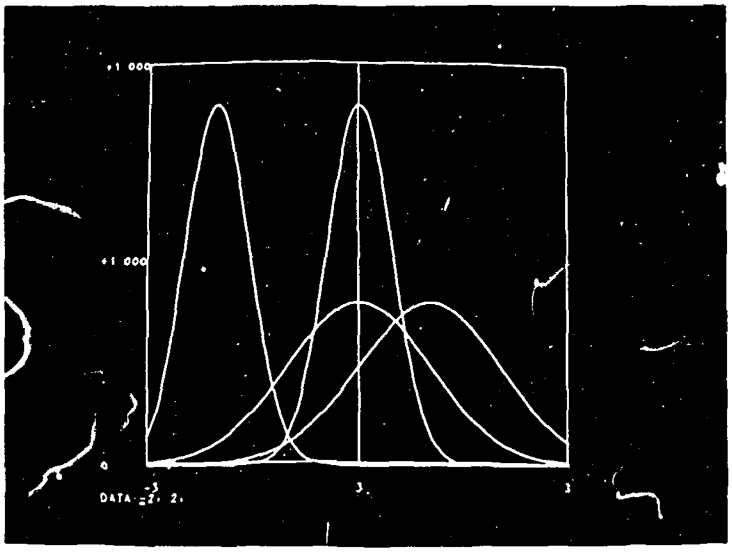

Figure 5 a) Display Corresponding to Exercise 1.

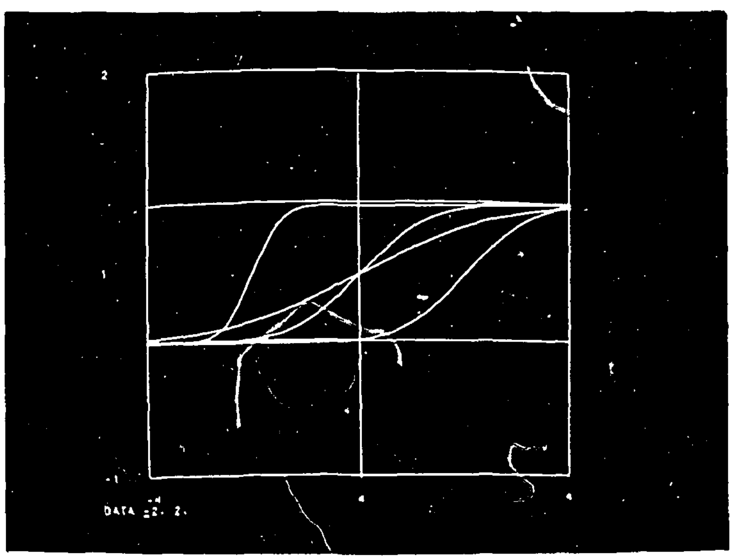

Figure 5 b) Display Corresponding to Exercise 2.

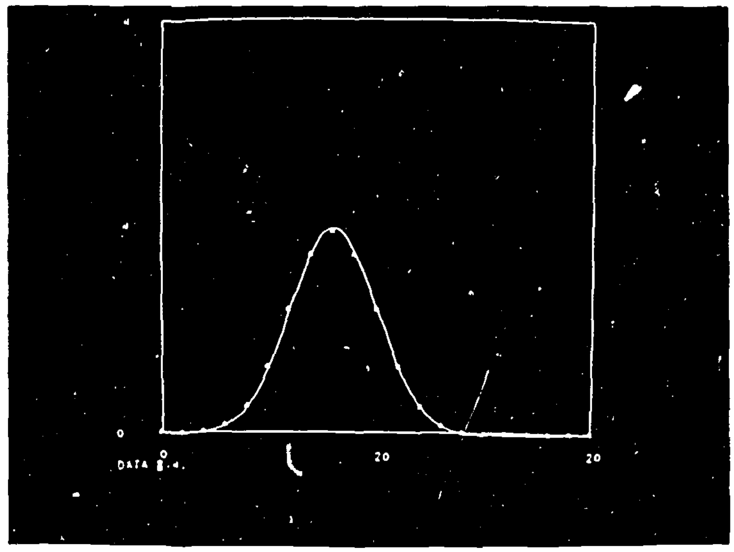

Figure 5 c) Display Corresponding to Exercise 3.

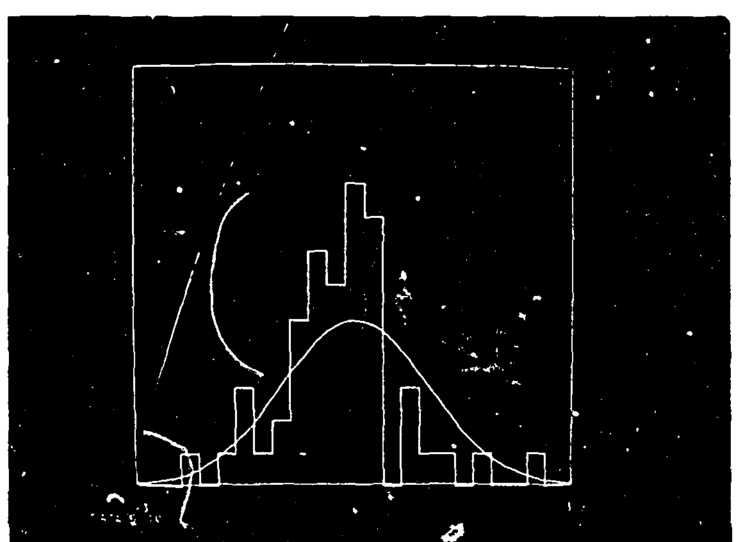

Figure 5 d) Display Corresponding to Exercise 4.

calculation. For this he must either remember the formulae for mean and variance of the binomial distribution or have some way of looking them up.

Next the student is asked to repeat the whole exercise but this time for cumulative distribution functions.

Exercise 4 is concerned with sampling. Several samples are drawn from the normal distribution and from the uniform distribution. The frequency distribution of the normalized means of the samples is displayed. These histograms are all superposed over a unit normal density function so as to allow for visual comparison with this curve.

No background in statistics is required to be able to perform these exercises mechanically, except for the calculation in Exercise 3. One would expect to derive maximum benefit from laboratory sessions shortly following classroom presentation of the subject matter. Clearly, the point in time for this varies from course to course even among introductory courses. For most courses of this type a good time would probably be around the seventh or eighth week. Perhaps Exercise 4 should be held off a few more weeks.

Exercise 1

Action		Comment
1.	RESET	pfk
2.	GRID	pfk
3.	Data:0,1,	enter distribution parameters.
4.	NORMAL	pfk
5.	Data:1,1,	enter distribution parameters.
6.	NORMAL	pfk; note that the change in the mean of the normal distribution corresponds to a <u>shift</u>.
7.	Data:1,2,	enter distribution parameters.
8.	NORMAL	pfk; a change in the variance corresponds to a change in the <u>spread</u> of the curve.
9.	REML	pfk; remove the last distribution plotted.
10.	Y.2	pfk; double the Y grid parameters.
11.	GRID	pfk; new grid.
12.	REDRAW	pfk; redraw everything on the new scales.
13.	Data:0,.2,	enter distribution parameters.
14.	NORMAL	pfk; a variance less than 1 makes the curve skinnier (and taller!) than N(0,1).
15.	Data:-2,.2,	enter distribution parameters.
16.	NORMAL	pfk.

*   *   *

Exercise 2

Action		Comment
1.	RESET	pfk
2.	xlo = -4 xhi = 4 xinc = 4 ylo = - 1 yhi = 2 yinc = 1	grid parameters.
3.	GRID	pfk
4.	CDF	pfk; cumulative distribution mode.
5.	Data:0,1,	enter distribution parameters.
6.	NORMAL	pfk; cdf of N(0,1)
7.	Data:2,1,	enter distribution parameters.
8.	NORMAL	pfk; a change in the mean constitutes a shift.
9.	Data:0,4,	enter distribution parameters.
10.	NORMAL	pfk; the curve is straighter than the previous ones.
11.	Data:-2,.2,	enter distribution parameters.
12.	NORMAL	pfk; notice the steep slope of the curve.

* * *

Exercise 3

Action		Comment
1.	RESET	pfk
2.	xlo = 0 xhi = 20 xinc = 20 ylo = 0 yhi = .4 yinc = .4	grid parameters
3.	GRID	pfk
4.	Data:4,.5,	enter distribution parameters.
5.	BINOM	pfk; binomial distribution.
6.	Data:5,.5,	enter distribution parameters.
7.	BINOM	pfk
8.	REM NL	pfk; remove the next to last distribution.
9.	Data:6,.5,	enter distribution parameters.
10.	BINOM	pfk
11.	REM NL	pfk; remove the next to last distribution.
12.	Data:12,.5,	enter distribution parameters.
13.	BINOM	pfk
14.	REM NL	pfk; remove the next to last distribution.
15.	Data:16,.5,	enter distribution parameters.
16.	BINOM	pfk
17.	REM NL	pfk; remove the next to last distribution.

The discrete points now look pretty much like a normal distribution. Which normal distribution? Try finding it by trial and error. Check your result by calculation.

Introduce new grid parameters and see what the correspondence is for the respective cumulative distribution functions.

* * *

Exercise 4

Action		Comment
1.	RESET	pfk
2.	xlo = - 3 xhi = 3 xinc = 6 ylo = 0 yhi = 1 yinc = 1	grid parameters
3.	GRID	pfk
4.	Data:0,1,	enter distribution parameters.
5.	NORMAL	pfk; pdf of N(0,1).
6.	Data:5,50,.5,	enter sampling parameters.
7.	SAMPLE	pfk; the system goes into the sample mode. sample size = 5 number of samples = 50 interval for histogram = .5
8.	Data:0,1,	enter distribution parameters.
9.	NORMAL	pfk; sample from N(0,1).
10.	REM L	pfk; remove the last distribution.
11.	Data:5,3,	enter distribution parameters.
12.	NORMAL	pfk; sample from N(5,3).
13.	REM L	pfk; remove the last distribution.
14.	Data:0,1,	enter distribution parameters.
15.	UNIF	pfk (16); sample from a continuous uniform distribution (0,1).
16.	REM L	pfk; remove the last distribution.
17.	Data:5,50,.25,	enter new sampling parameters.
18.	SAMPLE	pfk
19.	Data:5,10,	enter distribution parameters.
20.	UNIF	pfk (16); sample from a continuous uniform distribution (5,10).

* * *

3.4. Brief Description of programmed function keys (pfk's)

Number	Label	Description
00	INS1	Displays the general instructions
01	RESET	Displays the permanent display
02	GRID	Displays the grid
03	REDRAW	Redraws everything according to new scales
04	PDF	The system enters the probability density mode
05	REGEN	Regenerates displays
06	X/2	Divides the X grid parameters by 2
07	Y/2	Divides the Y grid parameters by 2
08	X.2	Multiplies the X grid parameters by 2
09	Y.2	Multiplies the Y grid parameters by 2
10	CDF	The system enters the cumulative distribution mode
11	SAMPLE	The system enters the sampling mode
12	UNIF	Displays the discrete uniform distribution
13	BINOM	Displays the binomial distribution
14	POISS	Displays the Poisson distribution
15	HYPER	Displays the hypergeometric distribution
16	UNIF	Displays the continuous uniform distribution
17	TRIAN	Displays the triangular distribution
18	NORMAL	Displays the normal distribution
19	GAMMA	Displays the gamma distribution

Number	Label	Description
20	EXPO	Displays the exponential distribution
21	CAUCHY	Displays the Cauchy distribution
22	BETA	Displays the beta distribution
23	STUD T	Displays Student's t distribution
24	CHI 2	Displays the chi square distribution
25	F	Displays the F distribution
26		Not implemented
27		Not implemented
28		Not implemented
29	REM NL	Removes the next to last displayed distribution
30	REML	Removes the last displayed distribution
31	END	Ends the program

3.5. Error Messages and User Response

Message	Meaning	User Response
FUNCTION PARMS ERROR	An error has been detected in the distribution parameters	Retype the parameters and press the distribution pfk again
GRID PARMS ERROR	An error has been detected in the grid parameters	Retype the parameter(s) and press GRID
BUFF OVERFLOW	An attempt has been made to plot more than seven distributions	Remove one or more distributions and try again
FUNCT-GRID INCOMPAT	The desired distribution does not exist within the current grid	Obtain a new grid and retry
SORRY:UNERASABLE ITEM	The user is attempting to erase something which is not a distribution	
SAMPLE PARMS ERROR	The user has entered incorrect sample parameters	Correct the parameters and press SAMPLE again
INPUT OUTPUT ERROR	The system is experiencing hardware trouble	If this persists notify the operator
NOT IMPLEMENTED	Sampling is not implemented for this distribution	

3.6. Glossary of Terms

Alphanumeric keyboard	- a typewriter-like device from which data may be entered onto the display area.
Programmed function keyboard	- keyboard containing 32 pushbutton keys labeled INST, RESET, etc.
pfk (programmed function key)	- one of the keys on the programmed function keyboard
cursor	- bright underscore on the screen indicating where data entered from the alphanumeric keyboard will appear
lightpen	- a pen-like device used to point at a distribution in order to erase that distribution
foot switch	- a switch in the form of a pedal which activates the lightpen
DATA field	- an area on the screen where distribution parameters are entered. Immediately to the right of the word DATA:

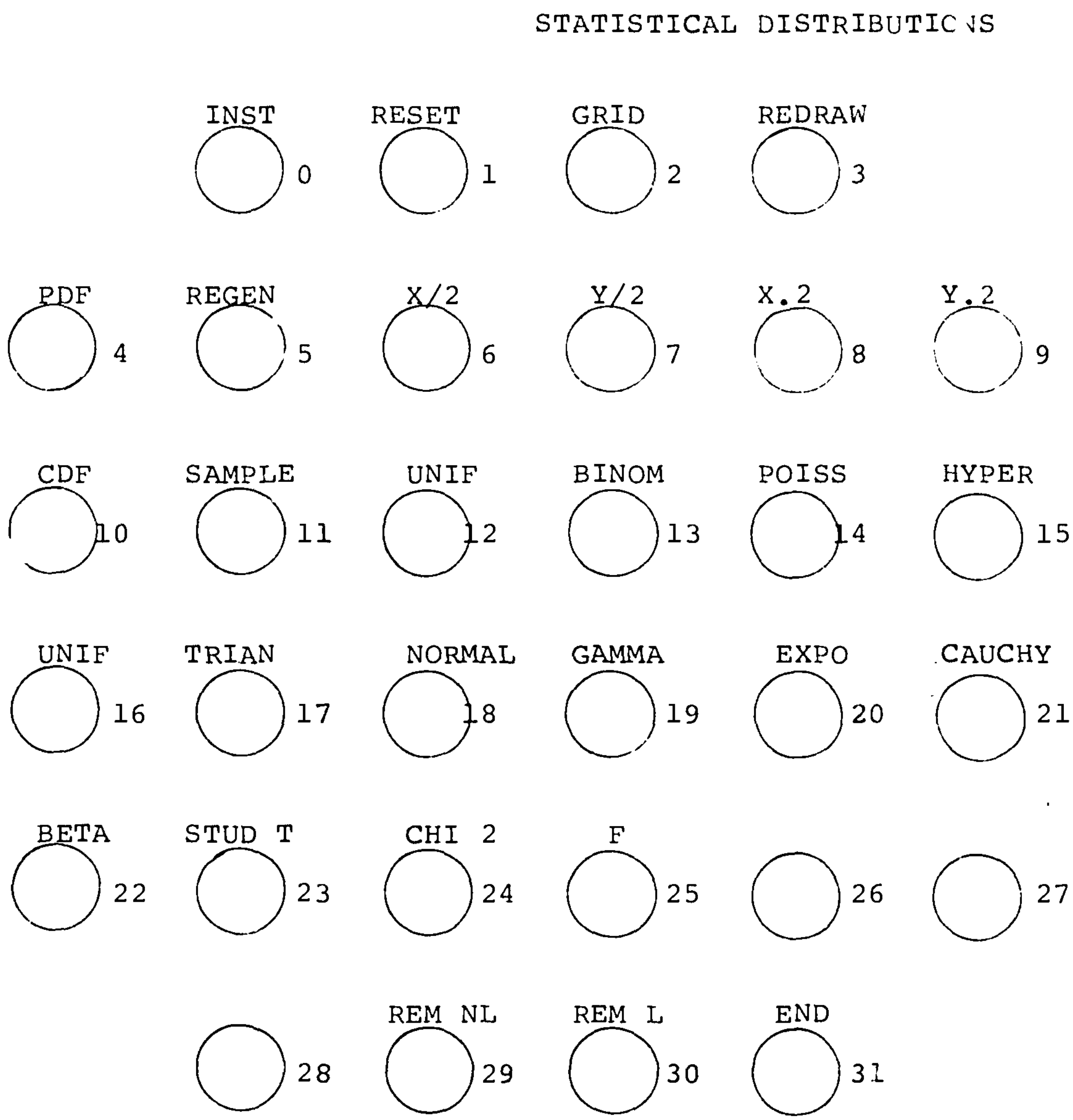

Figure 6. <u>The Programmed Function Keyboard</u>

1. Alphanumeric keyboard
2. Programmed function keyboard
3. pfk (programmed function key)
4. cursor
5. lightpen
6. foot switch
7. DATA field

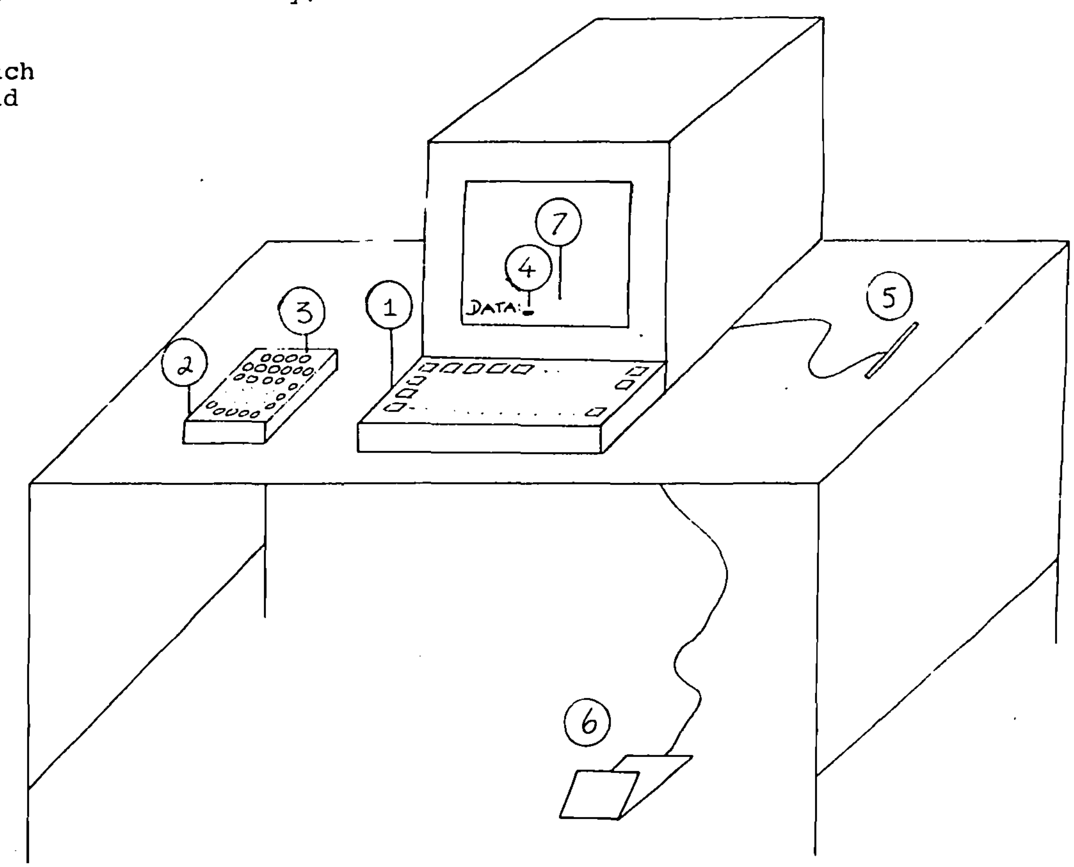

Figure 7. Sketch of Display System

IV. THE COMPUTER PROGRAM

4.1. Main Graphics Program

The statistical display system consists of 26 load modules written in Basic Assembler Language System/360, occupying a total of approximately 30K bytes of memory. All load modules reside in core at all times while the system is in operation. Figure 8 shows the logical structure. The arrow in that diagram indicates a calling sequence between two subprograms. For example, SAMPLER calls RAN; RAN calls no other program. Where boxes were drawn contiguously this denotes similarity of purpose among two or more subprograms; e.g., COP calls all the subprograms from BETA to STUDENT, all of which correspond to statistical distributions. The intermodular links are all according to OS/360 conventions except for the case of the link between COP and the subprograms called by it. In this instance the method of calling was done by creating a separate module (VCONS) consisting solely of a list of external address constants. COP, having received from ST2250 the number pertaining to the particular distribution to be called, extracts the address of the relevant subprogram from VCONS. This type of link was constructed in order to

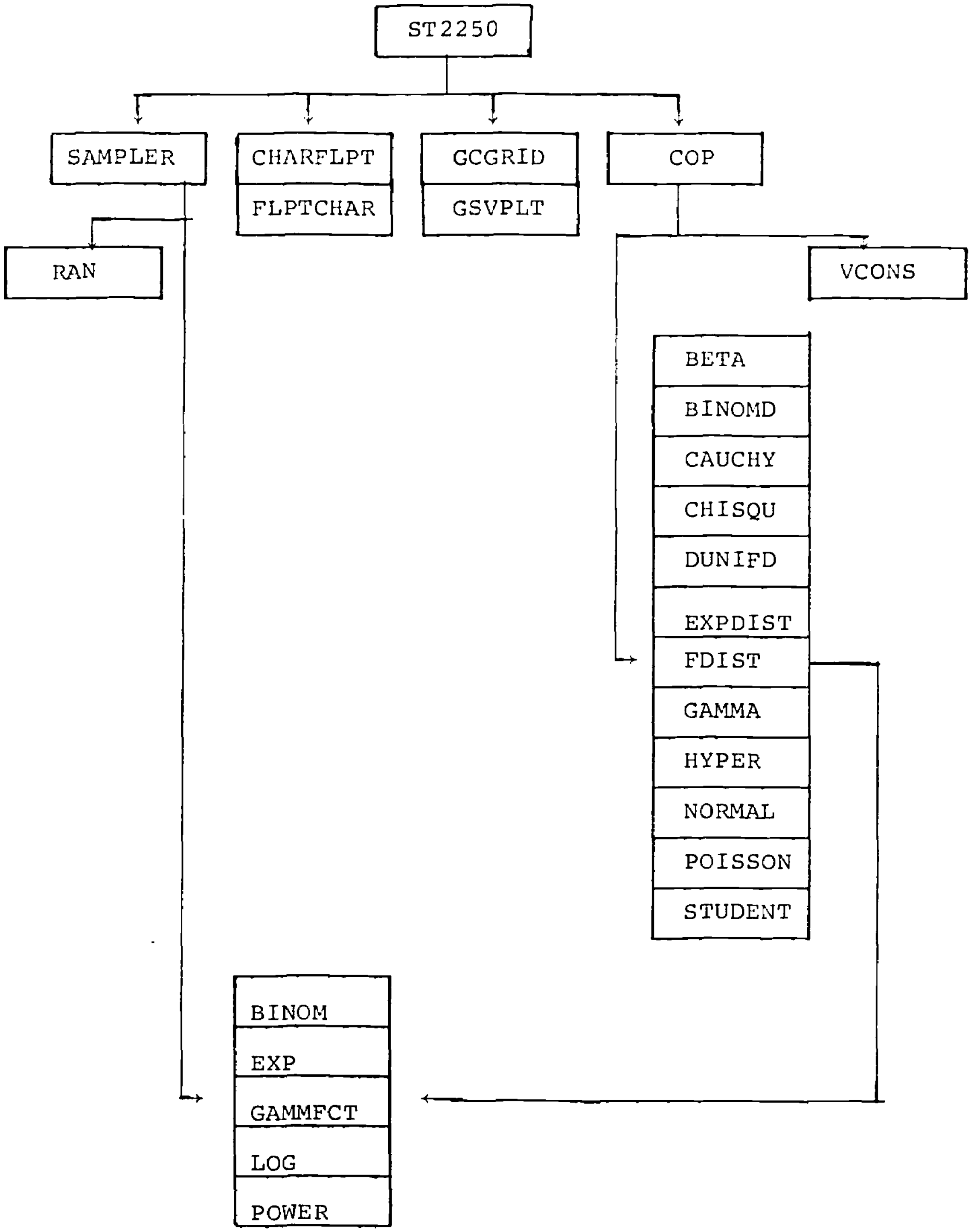

Figure 8. Logical Program Structure

facilitate the addition of new subprograms (pertaining to new statistical distributions). It was particularly helpful while building the system; new distributions could be added without having to recompile COP.

The main graphics program named ST2250 performs all input/output tasks. The most basic task is that of identifying attentions. Attentions occur whenever the user presses one of the pfk's or operates the lightpen. In the event that a pfk is pressed a further analysis has to be performed to identify the number of the particular pfk. ST2250 thus performs the following actions:

(1) requests 2250 buffer space;

(2) specifies attention routines; i.e., which attentions should be recognized by the system;

(3) activates (unlocks) the alphanumeric keyboard and the pfk lights;

(4) analyzes attentions;

(5) reads in the parameters entered via the alphanumeric keyboard;

(6) generates and deletes displays;

(7) calls the appropriate subroutines;

(8) delivers error messages;

(9) returns buffer storage;

(10) terminates the system.

Actions (1), (2), and (3) are done when the system is initialized; (9) and (10) are done when the system is terminated; (4) through (8) are done while the system is in operation.

The subprograms called directly by ST2250 are:

(1) SAMPLER, a module which draws pseudo-random samples from the various distributions and prepares a frequency distribution which is used by ST2250 to draw the histogram;

(2) FLPTCHAR, a subroutine to convert numbers from the floating point format to a character string;

(3) CHARFLPT, a subroutine to convert numbers from a character string to the floating point format;

(4) and (5) GCGRID and GSVPLT, generalized routines which generate graphic orders and data for displaying grids and plots respectively.

(6) COP, a service module which operates as a link between ST2250 and the statistical distribution routines.

All computations are done in short floating point (E) format. Numbers read in via the alphanumeric keyboard are in character format. Also any number to be displayed is required to be in character format. Thus FLPTCHAR and CHARFLPT are necessary in order to make these conversions. CHARFLPT detects the presence of non-numeric characters but interprets blanks as zeros.

SAMPLER calls RAN which generates pseudo-random numbers on the interval [0,1]. RAN is an additive congruential random number generator (see Reference 6). The basic idea is as follows: let r_i $i=1,2,\ldots,k$ be k random numbers such that $0 < r_i < 1$, all i. A new random number $r_{k+1} = r_1 + r_k$ (mod 1) is calculated. Then r_i' i=1,2,...,k are formed by $r_i' = r_{i+1}$. RAN does this with k equal to 18. The numbers are all floating point numbers.

The choice of method of pseudo-random number generation is not critical for this system. The most important consideration is speed of response. Some auxiliary programs were used to check the results of RAN for uniformity. This was done only by inspection. Because of the modularity of the system it is not a difficult matter to see if any other method of generation yields faster results. The current speed of response is quite satisfactory, however.

COP performs a wide variety of important tasks. All the curves to be displayed are produced by calculating a number of closely located points and joining these by linear segments. The number of points depends on the range of the curve. If the curve ranges over the entire grid 180 points are used; other ranges are prorated. ST2250 calculates how many points are required and passes on this information to COP along with a number identifying the required distribution. This number equals that of the corresponding pfk. Also passed to COP by ST2250 are the starting X value, the length of the X intervals, and an indication as to whether a pdf or a cdf is required. Samples are handled entirely through SAMPLE. On the basis of the information received COP accesses the right statistical distribution the necessary number of times. Each call to one of the statistical distribution routines yields one function value. In the event that a cdf is required COP performs a numerical integration by Simpson's Rule where h equals half the length of the X intervals. COP returns a vector of function (pdf or cdf) values. These values are used by ST2250 to plot the desired curve; i.e., ST2250 passes them on to GSVPLT.

4.2. Other Modules

4.2.1. Basic Functions

The following modules are used for calculating basic mathematical functions:

BINOM, which computes $\binom{n}{x}$;

EXP, which computes e^x;

GAMMFCT, which computes $\Gamma(x)$;

LOG, which computes $\log(x)$;

POWER, which computes x^a.

POWER with $a = .5$, rather than a separate square root routine, is used to extract square roots. All these functions require short floating point arguments and return short floating point values.

4.2.2. Statistical Distributions

The modules for computing the statistical distributions are:

BETA, beta distribution;

BINOMD, binomial distribution;

CAUCHY, Cauchy distribution;

CHISQU, chi square distribution;

DUNIFD, discrete uniform distribution;

EXPDIST, exponential distribution;

FDIST, F distribution;

GAMMA, gamma distribution;

HYPER, hypergeometric distribution;

NORMAL, normal distribution;

POISSON, Poisson distribution;

STUDENT, Student's t distribution.

These modules compute the value of the probability density function for a given argument. A call to any of these modules yields one answer. No special approximations are used except as incorporated in the basic mathematical functions used, if any. The particular form used of these distributions is recorded in section 3.2.3.

The following modules were not written by the author of this thesis: CHARFLPT and FLPTCHAR, which were written by Mr. D. Saville; EXP and LOG, which were written by Dr. A. Oliver; GCGRID and GSVPLT, which are part of the IBM standard graphics package.

V. SUMMARY

5.1. Extensions

The system was designed as a teaching tool. As such the foremost consideration was ease of use. Other considerations were speed of response and, of course, reliability. As is probably the case with most systems of this type, some features remain which one might have liked to be included. Among these are:

(1) The ability to enter and associate with a pfk an arbitrary distribution. This could be used to verify the feasibility of some approximation.

(2) The ability to display distributions of functions of random variables. In particular the ability to display the sum of two random variables might be a useful feature.

(3) The ability to work with live data, e.g., fitting raw data to some theoretical distribution.

(4) Some regression analysis, possibly in connection with (3).

5.2. Qualitative Measurement

The set of exercises contained in section 3.3 was tried on five students in an introductory statistics course (STAT

100; UNC; fall 1969). This course, being a service course for non-statistics majors, generally has students with widely varying backgrounds. As participation was put on a voluntary basis, one is led to suspect that the five students who actually participated (of the eight who originally signed up) tended to be among the better students, or in any case among the more interested. No quantitative measurement was attempted. There were three sessions of one hour each. Two sessions had two students and one session had one student. In each case the students performed the exercises themselves after brief verbal instructions. They were supervised during the entire session. In all cases the reactions of the students were favorable. They were able to go through the exercises without much difficulty. They seemed to be quite interested not only in the subject matter but also in the computer programming aspects of the system. The most frequently asked question concerned the possibility of calculating with live data.

Proof of effectiveness of a system of this type should, of course, be by statistical analysis of a controlled experiment. Isolating and quantifying the benefits to be reaped from such a system is not an easy task. Controlling for the novelty effect, in particular, would require a very careful design. It is not always clear that one needs to control for this effect, i.e., if a student being introduced to statistics derives extra motivation from a novel piece of equipment, so much the better. This is an oversimplification, of course, since the question remains as to effectiveness after the novelty has worn off.

An interesting experiment would be the measurement of the difference in effectiveness, if any, between the situations described in section 1.1., i.e., classroom versus laboratory situation. Ideally, one should also be able to determine experimentally the optimum number of students to have at one time in the laboratory situation. A matter which each course organizer would have to determine for himself is whether the expected benefits from having graphical display sessions are great enough to justify the cost involved. A compromise would seem to be either a film or a series of slides taken from the most interesting displays. In such a system the freedom to move in any direction, as prompted by questions from the students, is lost. A fruitful area of research might also be the measurement of the difference in effectiveness, if any, between such a film or slide system and a pure interactive system.

5.3. Justification of Graphical Display Use

The advantages of an interactive graphical display system to illustrate statistical distributions over, say, a book containing drawings of a representative number of statistical distributions are, among others:

(1) Continuous variation of parameters, thus providing an essentially infinite number of display possibilities.

(2) Superposition; this capability seems to be essential in order to compare adequately the shapes of distributions.

(3) Zooming capability; a system where one can freely vary extent and unit size of abcissa and ordinate is almost surely superior to a static collection of pictures.

(4) Not the least of added student motivation seems to derive from the knowledge that there is actually a conversation with a computer being held.

The effectiveness of an interactive graphical display system to aid the teaching of certain aspects of numerical analysis has been shown. (Reference 7).

BIBLIOGRAPHY

1. The Chemical Rubber Co., Standard Mathematical Tables, 15th Edition. Cleveland, Ohio, 1968.

2. Freund, J.E., Mathematical Statistics. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1962.

3. Hastings, C., Jr., Approximations for Digital Computers. Princeton, N.J.; The Princeton University Press, 1955.

4. International Business Machines Corporation. IBM System/360 Component Description IBM 2250 Display Unit Model 1, Form A27-22701, 1967.

5. International Business Machines Corporation. IBM System/360 Operating System Graphics Programming Services for IBM 2250 Display Unit, Form C27-6909, 1967.

6. Naylor, T.H.; Balintfy, J.L.; Burdick, D.S.; Chu, K., Computer Simulation Techniques. New York: John Wiley & Sons, Inc., 1966.

7. Oliver, A., "A Measurement of the Effectiveness of an Interactive Display System in Teaching Numerical Analysis." Doctoral Dissertation: University of North Carolina at Chapel Hill, 1969.

8. Wilks, S. S., Mathematical Statistics. New York: John Wiley & Sons, Inc., 1963.