A Computer-Based Laboratory for Mathematical Statistics and Probability

Elliot A. Tanis
Hope College
Holland, Michigan 49423
(616) 392-5111 Ext. 2267

Introduction

Hope College is a four-year, undergraduate, liberal arts college of 2,100 students. Most of the students who major in mathematics or computer science and some students who major in one of the natural sciences, psychology, or economics take a year-long, calculus-based course in mathematical statistics and probability.

During the last six years we have developed an optional computer-based laboratory that a student may elect to take in conjunction with the statistics course. Earlier papers [2,3] described the development of the laboratory and gave an outline of the topics that were covered. At that time Hope College owned an IBM 1130 computer. We now own a Xerox Sigma VI computer which has timesharing capabilities.

In this paper a summary of the earlier papers is given. In addition, recent developments and the availability of the materials is discussed.

Background and Description of Laboratory

We began to develop the laboratory materials and supporting computer software during the summer of 1971. It was motivated by students who had taken our statistics course and were interested in using the computer for research projects. The projects often required some sort of simulation. In the process of writing a computer program to properly perform a simulation, the student not only had to understand the underlying theory, he also gained a greater appreciation for this theory. We believed that similar types of experiences for more of our students would be beneficial. This prompted the development of the laboratory.

Originally the laboratory was mandatory. It is now optional because (1) some students had scheduling conflicts, and (2) some students did not like a computer-based laboratory for various reasons.

The laboratory meets for a two-hour block each week for an extra hour of credit. The first hour is spent discussing the previous week's assignment and introducing the new assignment. During the second hour the students begin to write their programs.

It is now possible for students to run their programs through batch processing or interactively at a terminal.

Laboratory Manual and Subprograms

An instructional package has been developed for a year-long laboratory for a mathematical statistics and probability course. The package includes a laboratory manual and supporting computer software.

The laboratory manual has gone through three revisions. It now contains more than 250 problems with complete solutions for twenty-five of them. The manual parallels the textbook *Probability and Statistical Inference* [1]. It can also be used with most of the standard textbooks.

The students write their own programs for solving the exercises. However over fifty subprograms are provided for their use. Among these are:

1. A pseudo-random number generator.
2. Graphing routines for: (1) the empirical distribution function with the option of

superimposing the theoretical distribution function, (2) a relative frequency ogive curve with the option of superimposing the theoretical distribution function, (3) a relative frequency histogram with the option of superimposing the theoretical probability density function, (4) a scatter diagram with the option of adding the least squares regression line, (5) a relative frequency power function with the option of superimposing the theoretical power function, and (6) printing the graph of any nonnegative valued function.

3. Programs that give the values of seven distribution functions.

4. Programs that give the values of the inverses of four distribution functions.

5. Programs that select random samples from the standard distributions.

It is possible to use all of these programs interactively at a terminal. The graphing routines use essentially a 100 x 50 character grid when run through batch processing and a 60 x 30 character grid when run interactively.

Goals

Some of the goals of the laboratory are to:

1. Help the student understand and gain a greater appreciation of the theory.
2. Illustrate that certain probability models are appropriate for particular random experiments.
3. Show how the computer can be used as an aid for analyzing data.
4. Expand possibilities for undergraduate research projects.

Examples

In this section two problems with complete solutions are given. These two problems were run using the interactive versions of RFHPF and CONIN. The subroutine RFHPF provides a Relative Frequency Histogram with the Probability Density Function superimposed. The subroutine CONIN depicts thirty CONfidence INtervals.

Problem 1

Let X_1, X_2, X_3 be a random sample of size 3 from a uniform distribution U(0,10). Thus $f(x) = 1/10$ and $F(x) = x/10$ for $0 < x < 10$.

(a) Define $g_1(y)$, $g_2(y)$, and $g_3(y)$, the marginal p.d.f.'s of the 3 order statistics, $Y_1 < Y_2 < Y_3$.

(b) Generate 200 samples of size 3, say x_1,x_2,x_3, from the uniform distribution U(0,10). For each sample let $y_1 = \min(x_1,x_2,x_3)$, $y_2 =$ median of (x_1,x_2,x_3), and $y_3 = \max(x_1,x_2,x_3)$.

(c) Use RFHPF to plot a histogram of the 200 observations of Y_1 along with $g_1(y)$, find the sample mean and sample variance of these observations.

(d) Use RFHPF to plot a histogram of the 200 observations of Y_2 along with $g_2(y)$. Find the sample mean and sample variance of these observations.

(e) Use RFHPF to plot a histogram of the 200 observations Y_3 along with $g_3(y)$. Find the sample mean and sample variance of these observations.

Problem 2

Let X be N(16.4,2.9).

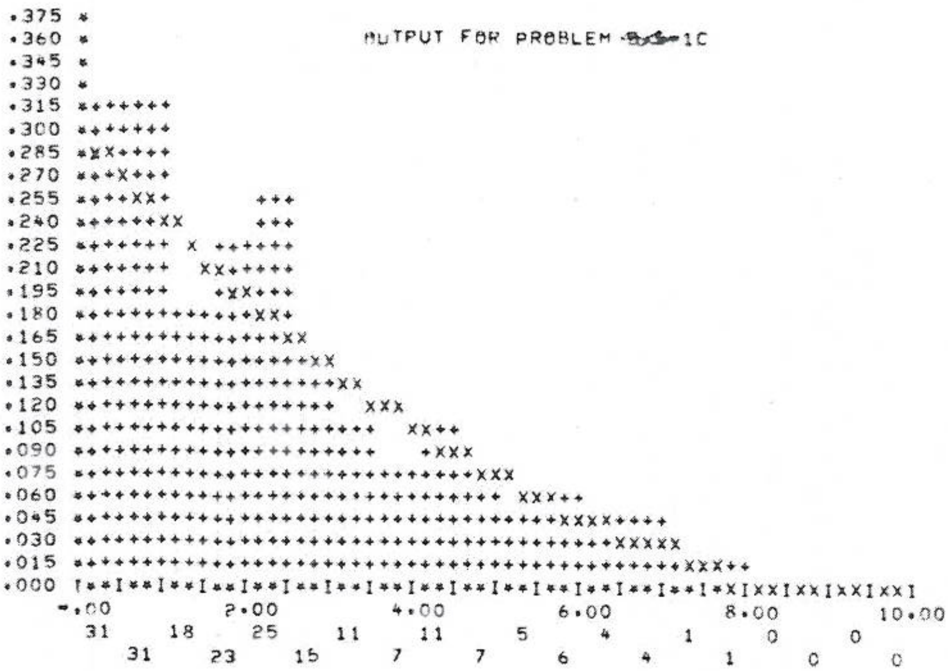
(a) Generate 50 samples of size n = 16 from this distribution. For each sample, find the values of Y_5 and Y_{12}, the 5th and 12th order statistics. Store these values in two arrays, say Y5 and Y12, which have been dimensioned to 50. (Use 30 for on-line version.) Hint: You may use the subroutine ORDER.

(b) Verify that (y_5,y_{12}) is a 92.32% confidence interval for $\pi_{.5} = 16.4 = \mu$.

(c) Approximately 92.32% of the 50(30) confidence intervals from part a should contain the median. To illustrate this, use the subroutine CONIN(14.0,18.8,Y5,Y12,16.4) to depict the confidence intervals.

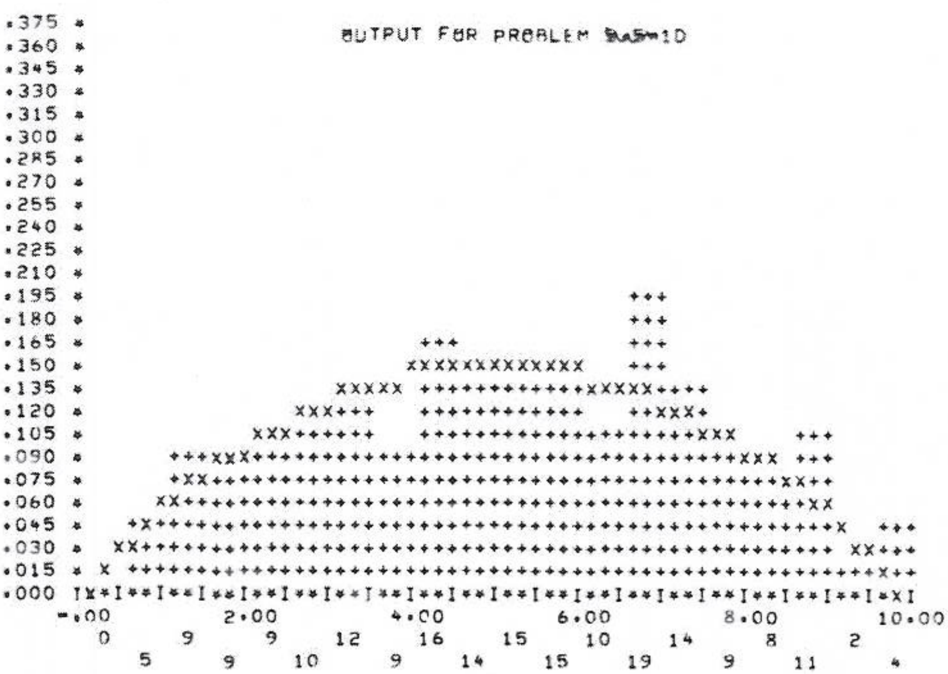
(d) What proportion of the confidence intervals contain the median?

```
C . . . SOLUTION FOR PROBLEM [redacted]1.
C
C . . . DIMENSION X TO HOLD EACH SAMPLE OF SIZE 3, Y TO HOLD THE
C . . . ORDERED SAMPLE OF SIZE 3, Y1,Y2,Y3 TO HOLD THE 200
C . . . OBSERVATIONS OF EACH ORDER STATISTIC.
C . . . DEFINE G1,G2,G3, THE P.D.F.'S OF THE THREE ORDER STATISTICS
C . . . Y1,Y2,Y3, RESPECTIVELY, AS STATEMENT FUNCTIONS.
C
      DIMENSION X(3), Y(3), Y1(200), Y2(200), Y3(200), NOBS(20)
      G1(U) = 3.*((1. - U/10.)**2)*(1./10.)
      G2(U) = 6.*(U/10.)*(1. - U/10.)*(1./10.)
      G3(U) = 3.*((U/10.)**2)*(1./10.)
C
C . . . GENERATE 200 OBSERVATIONS OF Y1,Y2,Y3.  FIRST GENERATE THREE
C . . . OBSERVATIONS OF X.  USE THE SUBROUTINE ORDER TO OBTAIN THE
C . . . VALUES OF THE ORDER STATISTICS.
C
      DO 20 K=1,200
      DO 10 J=1,3
      X(J) = 10.*RAN(1)
   10 CONTINUE
      CALL ORDER(3,X,Y)
      Y1(K) = Y(1)
      Y2(K) = Y(2)
      Y3(K) = Y(3)
   20 CONTINUE
C
C . . . PLOT HISTOGRAMS OF THE OBSERVATIONS OF Y1, THEN Y2,
C . . . THEN Y3, WITH THE RESPECTIVE P.D.F.'S SUPERIMPOSED. ALSO
C . . . CALCULATE, USING SMASV, AND PRINT THE RESPECTIVE SAMPLE
C . . . MEANS AND SAMPLE VARIANCES.
C
      WRITE(108,850)
  850 FORMAT('1',//,8X,'OUTPUT FOR PROBLEM 5.5-1C',//)
      CALL RFHPF(1,20,0.0,10.0,0.375,200,Y1,NOBS,G1)
      WRITE(108,800)
  800 FORMAT('0',8X,'HISTOGRAM AND PDF FOR THE SMALLEST ORDER STATISTIC'
     *,/)
      CALL SMASV(Y1BAR,Y1VAR,200,Y1)
      WRITE(108,825) Y1BAR,Y1VAR
  825 FORMAT(' ',8X,'THE SAMPLE MEAN IS',F8.3,/,8X,'THE SAMPLE VARIANCE '
     *IS',F8.3)
      WRITE(108,860)
  860 FORMAT('1',//,8X,'OUTPUT FOR PROBLEM 5.5-1D',//)
      CALL RFHPF(1,20,0.0,10.0,0.375,200,Y2,NOBS,G2)
      WRITE(108,805)
  805 FORMAT('0',8X,'HISTOGRAM AND PDF FOR THE MEDIAN',/)
      CALL SMASV(Y2BAR,Y2VAR,200,Y2)
      WRITE(108,825) Y2BAR,Y2VAR
      WRITE(108,870)
  870 FORMAT('1',//,8X,'OUTPUT FOR PROBLEM 5.5-1E',//)
      CALL RFHPF(1,20,0.0,10.0,0.375,200,Y3,NOBS,G3)
      WRITE(108,810)
  810 FORMAT('0',8X,'HISTOGRAM AND PDF FOR THE LARGEST ORDER STATISTIC'
     *,/)
      CALL SMASV(Y3BAR,Y3VAR,200,Y3)
      WRITE(108,825) Y3BAR,Y3VAR
      STOP
      END
```



HISTOGRAM AND PDF FOR THE SMALLEST ORDER STATISTIC

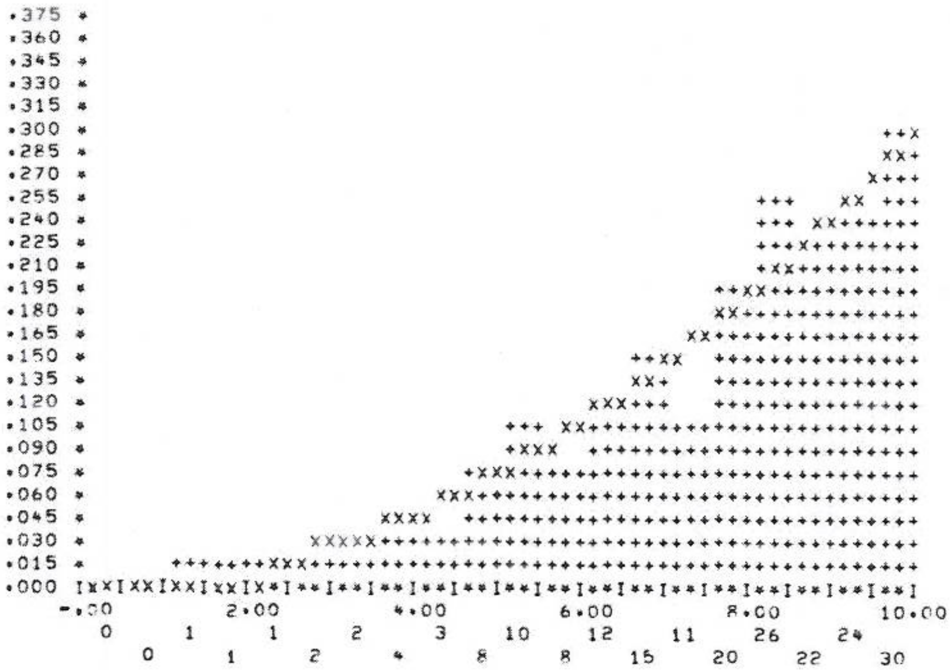
THE SAMPLE MEAN IS 2.320
THE SAMPLE VARIANCE IS 3.297



HISTOGRAM AND PDF FOR THE MEDIAN

THE SAMPLE MEAN IS 5.176
THE SAMPLE VARIANCE IS 5.341

OUTPUT FOR PROBLEM 5.5-1E



HISTOGRAM AND PDF FOR THE LARGEST ORDER STATISTIC

THE SAMPLE MEAN IS 7.554
THE SAMPLE VARIANCE IS 3.497

```
C . . . SOLUTION FOR PROBLEM
C
C . . . DIMENSION X TO HOLD THE ORIGINAL SAMPLE AND Y TO HOLD
C . . . THE ORDERED SAMPLE OF SIZE 16.  DIMENSION Y5 AND Y12
C . . . TO HOLD THE 5TH AND 12TH ORDER STATISTICS FOR THE 30
C . . . (50 IF USING BATCH VERSION) SAMPLES OF SIZE 16 FROM
C . . . THE NORMAL DISTRIBUTION N(16.4,2.9).
C
      DIMENSION X(16), Y(16), Y5(30), Y12(30)
C
C . . . LET M=30 FOR TIME SHARING, 50 FOR BATCH.
C
      M = 30
      DO 10 K=1,M
C
C . . . USE EITHER THE SUBROUTINE NORM OR THE FUNCTION
C . . . SUBPROGRAM XNRML TO GENERATE THE SAMPLE OF SIZE 16
C . . . FROM THE NORMAL DISTRIBUTION N(16.4,2.9).
C
      CALL NORM(16.4,2.9,16,X)
C
C . . . USE THE SUBROUTINE ORDER TO ORDER THE RANDOM SAMPLE.
C
      CALL ORDER(16,X,Y)
C
C . . . FILL ARRAYS Y5 AND Y12 WITH THE 5TH AND 12TH ORDER
C . . . STATISTICS.
C
      Y5(K) = Y(5)
      Y12(K) = Y(12)
   10 CONTINUE
C
C . . . USE THE SUBROUTINE CONIN TO DEPICT THE CONFIDENCE
C . . . INTERVALS.
C
      WRITE(108,850)
  850 FORMAT('1',8X,'OUTPUT FOR PROBLEM 6.1-1',//)
      CALL CONIN(14.0,18.8,Y5,Y12,16.4)
      WRITE(108,800)
  800 FORMAT('0',5X,'EACH INTERVAL IS A 92.32% CONFIDENCE INTERVAL')
C
C . . . COUNT THE NUMBER OF INTERVALS THAT CONTAIN THE
C . . . MEDIAN (MEAN) 16.4.
C
      NUM = 0
      DO 20 K=1,M
      IF(Y5(K).LE.16.4.AND.16.4.LE.Y12(K)) NUM = NUM + 1
   20 CONTINUE
      PROP = FLOAT(NUM)/FLOAT(M)
      PERCENT = 100.*PROP
      WRITE(108,805) PERCENT
  805 FORMAT('0',5X,F5.2,' PER CENT OF THE INTERVALS CONTAIN THE MEDIAN
     *')
      STOP
      END
```

```
OUTPUT FOR PROBLEM
```

```
CONFIDENCE INTERVALS
```


```
(15.107,17.815)
(15.543,16.727)
(15.117,17.567)
(16.452,18.414)
(15.303,17.614)
(15.439,17.289)
(15.200,17.597)
(15.665,17.012)
(15.298,17.205)
(15.238,17.718)
(14.954,17.146)
(15.534,17.563)
(15.176,17.291)
(14.783,17.272)
(15.171,17.343)
(14.579,16.479)
(15.332,17.536)
(15.580,17.606)
(15.389,18.282)
(15.917,18.221)
(15.552,17.820)
(15.695,16.809)
(15.301,17.776)
(14.958,17.814)
(15.761,17.641)
(15.657,16.681)
(16.145,17.561)
(15.411,17.269)
(15.266,16.600)
(16.483,17.007)
```

```
PARAMETER=  16.400

     EACH INTERVAL IS A 92.32% CONFIDENCE INTERVAL

     93.33 PER CENT OF THE INTERVALS CONTAIN THE MEDIAN
```

Conclusions and Recommendations

This laboratory provides a worthwhile experience for the student. The interactive capability is beneficial for class demonstration and it also allows a student to obtain immediate output. Some students using teletype terminals that print at ten cps have a tendency to tie up the terminal, especially when several graphs are required. Thus batch processing is still the best for some solutions and for some students.

Graphics terminals could have a positive impact on this laboratory. Unfortunately, Hope College does not yet own a graphics terminal, although proposals have been written which include the purchase of graphics terminals.

Distributions of Materials

The laboratory manual and subprograms have been favorably reviewed by CONDUIT. Copies of the laboratory manual and a tape containing the subprograms are available from CONDUIT, P.O. Box 388, Iowa City, Iowa 52240.

References

1. Robert V. Hogg and Elliot A. Tanis, *Probability and Statistical Inference*, Macmillan Publishing Co., Inc., New York, 1977.
2. Elliot A. Tanis, "Theory of Probability and Statistics Illustrated by the Computer," Proceedings of the 1972 Conference on Computers in Undergraduate Curricula, pp. 513-20 (1972).
3. Elliot A. Tanis, "A Computer Laboratory for Mathematical Probability and Statistics," Proceedings of a Fourth Conference on Computers in the Undergraduate Curricula, pp. 416-26 (1973).