



## Three Tools for Interactively Visualizing Some Distribution Theory Concepts

Peter K. Dunn

*The American Statistician*, Vol. 53, No. 2. (May, 1999), pp. 137-139.

Stable URL:

<http://links.jstor.org/sici?sici=0003-1305%28199905%2953%3A2%3C137%3ATTFFIVS%3E2.0.CO%3B2-0>

*The American Statistician* is currently published by American Statistical Association.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

# Three Tools for Interactively Visualizing Some Distribution Theory Concepts

Peter K. DUNN

In this article, three Matlab functions are introduced to help students visualize some concepts in distribution theory. The functions allow the students to experiment with the parameters and views to see the effects immediately. The three tools cover the areas of the central limit theorem, the normal approximation to the binomial, and the bivariate normal distribution. Some notes on using the programs in the classroom are also given.

**KEY WORDS:** Binomial; Bivariate; Central limit theorem; Gaussian; Matlab; Visualization.

## 1. INTRODUCTION

Many concepts in the theory of distributions can be difficult for students to visualize, yet are not difficult to demonstrate. The central limit theorem, for example, is fundamental to much in statistics yet is often poorly understood. As teachers, we may be happy to deal in concepts and theorems, but many students find this difficult and prefer concrete examples. In this article, three tools are introduced for use in Matlab for visually demonstrating some concepts of distribution theory in an interactive manner. In this interactive environment, students can observe the effects of the theory in action.

The programs are available free of charge from the MathWorks Worldwide Web site, at <http://www.mathworks.com/ftp/ftpindexv5.shtml>. Select the statistics section, and search for the three files: `cltexp.m`, `normbin.m`, and `bvnormal.m`. Once the files have been downloaded, they can be placed anywhere on the Matlab path and used as indicated below by typing the name of the function at the Matlab prompt. The Matlab path can be checked by typing `path` at the Matlab prompt. Adjusting the Matlab path is platform dependent, so typing `help path` at the Matlab prompt is recommended for instruction on this.

The functions require that Matlab version 5 be used (in any operating system for which Matlab is available), but the Statistics Toolbox is not necessary. In addition, students do not need to learn Matlab to use the functions, since they are used through a graphical user interface.

In the next section, the Matlab program for the central limit theorem is discussed. In Section 3, a program for

demonstrating the normal approximation to the binomial is presented; and in Section 4 a program for visualizing the bivariate normal distribution is discussed.

## 2. THE CENTRAL LIMIT THEOREM

### 2.1 The Program

Though a very important statistical result, the central limit theorem is often poorly understood by students. Despite this, it is relatively easy to demonstrate (even in a first-year class) by using a computer to choose random samples from a distribution, calculating means and plotting them. For a more experimental and interactive method—see, for example, Scheaffer, Gnanadesikan, Watkins, and Witmer (1996, p. 134). Here, we present a Matlab routine for demonstrating interactively that the sampling distribution tends to the normal even when the parent distribution is skewed, as in the case of an exponential distribution.

The program displays the sampling distribution from an exponential distribution with mean  $\beta$ , where  $\beta$  is allowed to be chosen such that  $.01 \leq \beta \leq 10$ . Samples can be chosen with sizes between 1 and 200. It is a simple exercise to show that the actual sampling distribution of samples of size  $n$  from an exponential distribution  $X \sim \exp(\beta)$  is given by  $\bar{X} \sim \text{Gam}(n, \beta/n)$  using, for example, moment generating functions.

### 2.2 Using the Program in the Classroom

The program is started in Matlab by typing `cltexp` at the Matlab prompt. This produces the screen shown in Figure 1. The bottom panel sets the parameters of the sampling

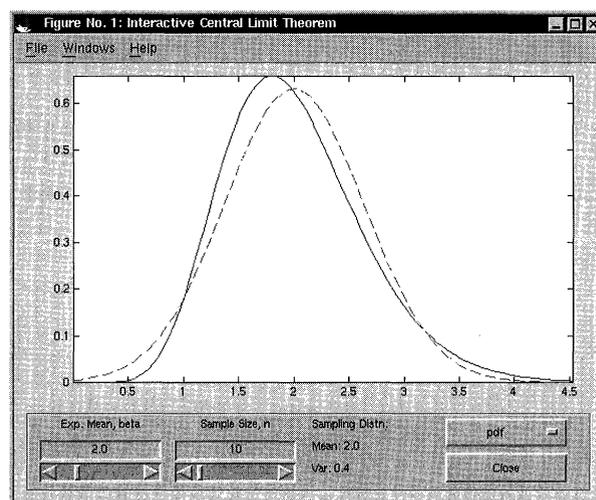


Figure 1. The Initial Screen for the Central Limit Theorem Program. By default, the sample size used is  $n = 10$  and the parent population is an exponential distribution with mean  $\beta = 2$ .

Peter K. Dunn is a Lecturer in the Department of Mathematics and Computing, University of Southern Queensland, Toowoomba, 4350, Australia (Email: [dunn@usq.edu.au](mailto:dunn@usq.edu.au)). The author wishes to thank the reviewers for their helpful comments and for pointing out a further relevant reference.

distribution: the mean of the parent exponential distribution, and the sample size. It also displays the mean and variance of the sampling distribution, and allows the user to choose the probability density function or the cumulative density function as the display. The Close button closes the session.

The parameters can be chosen by either (a) sliding the slide control (or moving it using the arrows) or (b) by entering a value in the edit box. In either case, the graph is redrawn immediately so the effect can be seen.

Using the program, students can see the tendency of the sampling distribution to approach the normal, even with a skewed parent population. It is an instructive exercise for the students to determine the exact sampling distribution using moment generating functions, and then see the tendency of the resulting distribution approach the normal. A slightly different but equally instructive approach to the central limit theorem is given in one of the Lisp-Stat programs discussed by Marasinghe, Meeker, Cook, and Shin (1996), where random samples are drawn from a specified distribution and displayed using histograms together with the parent distribution.

### 3. THE NORMAL APPROXIMATION TO THE BINOMIAL

#### 3.1 The Program

The program is started by typing `normbin` at the Matlab prompt, producing the initial screen shown in Figure 2. The bottom panel is similar to that in Figure 1. The parameters can be set so that the probability of success  $p$  is between .01 and .99, and the sample size  $n$  is between 3 and 100. The mean and the variance of the distribution is also displayed. Again, the parameters can be chosen by either (a) sliding the slide control (or moving it using the arrows) or (b) by entering a value in the edit box. In either case, the graph is redrawn immediately so the effect can be seen.

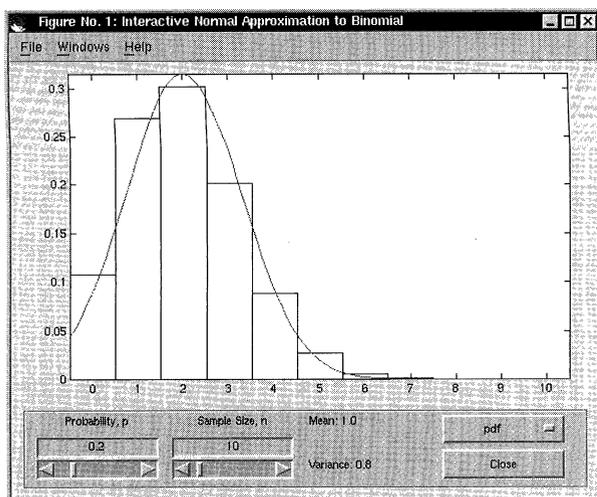


Figure 2. The Initial Screen for the Normal Approximation to the Binomial Program. By default, the sample size used is  $n = 10$  and the probability of a success is  $p = .2$ .

### 3.2 Using the Program in the Classroom

The students can alter the values of  $n$  and  $p$  to see the effect of changing the parameters. It is easy to see that for  $p$  near 0 or 1 that the approximation is quite poor, and conversely that the approximation can be adequate near  $p = .5$  for a moderate sample size.

The effect of the sample size can also be demonstrated. Even when  $p$  is not near .5, larger sample sizes produce binomial distributions close to normal. Most textbooks suggest that the normal approximation to the binomial is adequate if  $np$  and  $n(1 - p)$  both exceed a certain threshold, often 5 or 10. Students can examine why such guidelines are given by observing the distribution near these limits.

The necessity of the continuity correction is also clear from comparing the binomial and normal distributions. As the sample size increases, the need for the continuity correction disappears as the binomial distribution tends to the normal.

## 4. THE BIVARIATE NORMAL DISTRIBUTION

#### 4.1 The Program

The program is contained in the Matlab function `bvnormal.m`. The session is started by typing `bvnormal` at the Matlab prompt and the initial screen presented is shown in Figure 3. (The actual program uses colors; they are represented here in black and white using the Black and White option in the program.)

This program has a screen more complex than the previous two programs.

- The main screen shows the bivariate normal density as specified by the parameters at the bottom of the screen. The plot can be rotated by clicking and holding down the left mouse button, moving to a suitable vantage point, and releasing the mouse button.
- The parameters of the distribution are given at the bottom of the screen. The parameters can be changed in one of two ways: (a) by clicking on the arrows to increase or

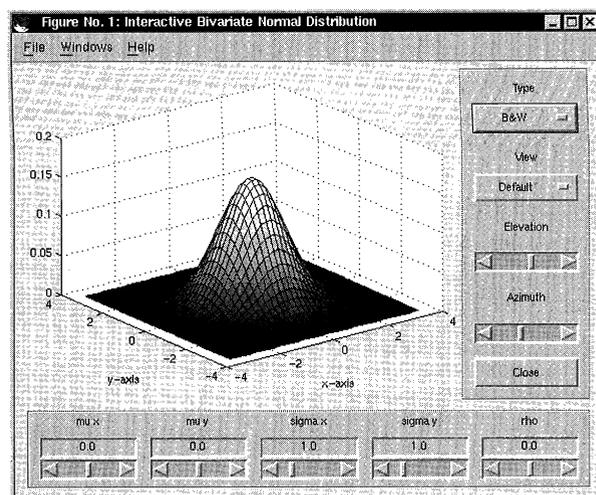


Figure 3. The Initial Screen. The plot is in the center, the viewing commands are on the right, and the parameter settings are along the bottom.

decrease the parameter value; or (b) by entering a particular value in the editable region just above the slider controls. The means can vary between  $-10$  and  $+10$ ; the standard deviations between  $.1$  and  $5.0$ , and  $\rho$  (the correlation coefficient) between  $-.99$  and  $.99$ . By default, the means are zero, the standard deviations one, and the correlation coefficient zero.

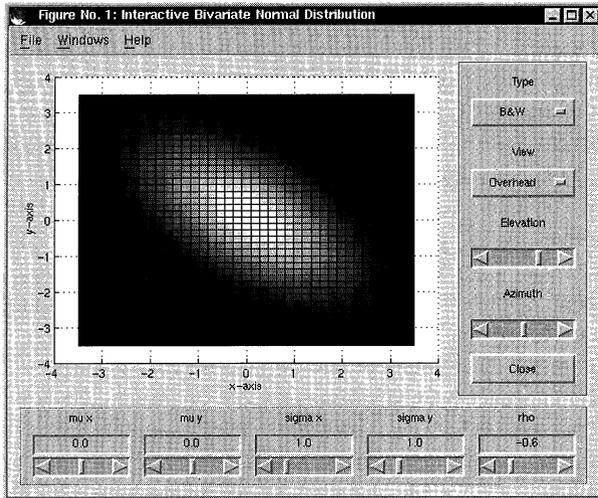


Figure 4. An Overhead View With  $\rho = -.6$ . The effect of changing  $\rho$  can be clearly seen in the overhead view.

- The view of the distribution is controlled by the panel on the right. The Type of display is either a mesh (the default), a surface, or black and white. The View can be altered to be an overhead view, a view parallel to either axis, or the default view as shown in Figure 1. The Elevation and Azimuth adjust the view the user has of the distribution. (Recall that these can also be changed interactively by clicking on the plot itself.) The Close button closes the session.

The choices under the View menu have been chosen to give instructive views of the distribution, which are discussed in the next subsection.

#### 4.2 Using the Program in the Classroom

Apart from adjusting the parameters in the bottom panel and seeing the effect, the program can be used in many instructive ways.

Often students have difficulty in understanding the effect of the correlation coefficient,  $\rho$ , and believing that the marginal distributions (or the conditional distributions at a given value of one of the variables) are normal. One way of demonstrating the effect of the correlation coefficient is to select the Overhead view from the View menu. The display shows an overhead shot of the distribution where the colors (or shades of gray) indicate the height of the proba-

bility density function. Adjusting the correlation coefficient produces the appropriate change in the distribution that becomes visible on the plot; see Figure 4, where  $\rho$  has been set at  $-.6$ . For a better true three-dimensional view, the Elevation can be decreased and the Azimuth altered a little. The resulting plot shows the linear relationship between the two variables and can be compared to a scatterplot of data with a similar  $\rho$  to show the connection.

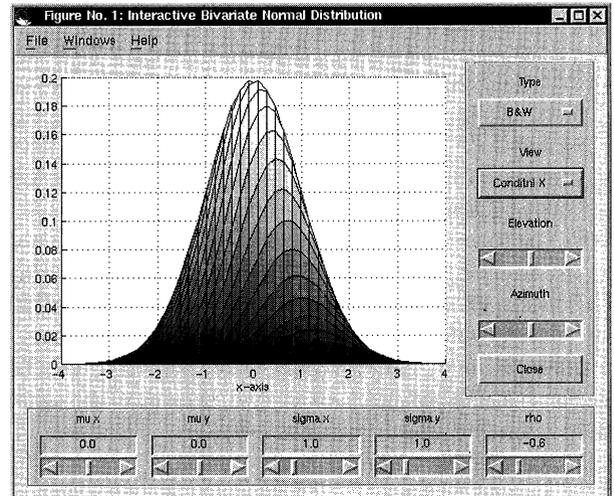


Figure 5. A View of the Conditional Distributions of  $X$  with  $\rho = -.6$ . The lines running from the left of the plot to the right indicate (nonnormalized) conditional distributions of  $X$  for different values of  $y$ .

The effect of  $\rho$  on the conditional distributions can be seen by selecting Conditi X (or Conditi Y). The plot is shown in Figure 5. (Again, a slight adjustment of the Elevation and Azimuth may prove to be useful.) The contours show that the (nonnormalized) conditional distributions of  $X$  for different  $y$ -values all have normal distributions. Changing the value of  $\rho$  is also constructive; the conditional distributions can be seen changing as  $\rho$  changes but remaining normal. In addition, when  $\rho = 0$ , it can be seen that the mean of the  $X$ -variable is the same for all values of  $y$ , but for  $\rho \neq 0$  the expected value of  $X$  depends of the given  $y$  value.

[Received June 1998. Revised September 1998.]

#### REFERENCES

- Marasinghe, M. G., Meeker, W. Q., Cook, D., and Shin, T.-S. (1996), "Using Graphics and Simulation to Teach Statistical Concepts," *The American Statistician*, 50, 342-351.
- The MathWorks, Inc. (1997), *The Matlab Reference Guide version 5*, Natick, MA: Author.
- Scheaffer, R. L., Gnanadesikan, M., Watkins, A., and Witmer, J. A. (1996), *Activity-Based Statistics Instructor Resources*, New York: Springer.