INTRODUCTION

It is often the case that graduate and undergraduate students emerge from their first statistics course convinced that the topic is boring and difficult. Many otherwise bright students seem to bog down in statistics courses. One reason for this phenomenon might be the way statistics courses are taught. Regardless of whether a practical, cookbook approach is used, or a more theoretical approach, the task of the student is to memorize or otherwise learn long strings of numerical text. There is a fair amount of literature in the field of cognitive psychology that would indicate many people learn and memorize better when the material to be learned can be visualized, or imaged, in the mind (see Paivio, 1971 for a review).

Statistical distributions can be looked at as either abstract algebraic entities, or as geometric objects. With that in mind, we have taken the equations for a number of common and not so common distributions, and attempted to create pictorial representations for these distributions. These will enable us to observe some of the features of the distributions. We will also be able to examine the relationships between some of these distributions. While many statistical texts contain two dimensional pictures of distributions, we have used SAS/GRAPH to create a number of three dimensional pictures. Three dimensional pictures are especially useful for representing bivariate relationships, and how the parameters of a distribution change its shape.

Distributions can be generated with ease using a SAS DATA step. Below we present an outline of how this is done.

```sas
DATA DIST ;
DO X = -2.5 TO 2.5 BY .1 ;
Y = some function of X ;
OUTPUT ;
END ;
```

In this example, the domain of the function goes from -2.5 to 2.5. The OUTPUT statement sends the variables X and Y to the SAS dataset DIST for later plotting. For three dimensional plots, there would be another DO-END loop nested within the first. The formulas for the functions themselves can be found in a number of advanced statistics texts, such as Theil (1971) and especially Johnson and Katz (1970).

Generation of distributions in this manner was greatly facilitated by the built-in functions provided in the SAS DATA step. For instance, the normal distribution function is available and is called PROBNORM. There are a number of mathematical functions which facilitate the calculation of distribution functions. In this category would be included the GAMMA function.

THE STATISTICAL DISTRIBUTIONS

It is important for students to know the difference between a density function and a cumulative distribution function. The univariate normal distribution is used to illustrate this difference in Figure 1. The notion of an integral need not be explained to students if they can acquire an intuitive grasp of what a tail probability is. As we sweep through the density function, the distribution function builds to a probability of 1.00.

The normal density function is described with two parameters. In Figure 2, the mean is varied along the x axis while everything else is held constant, including the variance, which is set to 1.000. In Figure 3, the variance is changed while the mean is held constant.

Now, consider the joint variability of two normal variates. The bivariate normal distribution can be described by five parameters. There are two means, two variances, and a covariance. In Figure 4, the two variables are standardized to a mean of zero and unit variance, and the covariance between them is set to .95. As is well known, the covariance between standardized variates is the correlation. Other correlations can be used to show how the joint density changes shape as the correlation changes. In the interest of space, only a .95 correlation is used here. The curves come from the purely mathematical definition of bivariate normal (Bock, 1975 p. 123). SAS has the capability of generating random normal deviates, and this was done to compare the mathematical form with a Monte Carlo style simulation. When 10,000 pairs of normal deviates are generated with the SAS NORVAL function, the result of the joint density is shown in Figure 5. The expected correlation in this figure would be 0. When 100,000 pairs are generated for Figure 6, the function becomes more smooth, and very nearly resembles the theoretical form.

One feature of these bivariate normal figures that could be pointed out to students is that the conditional densities for either of the two variables remains normal, regardless of the correlation. This property can be observed by...
looking at any individual curve parallel to either axis in Figure 4. Up to now, we have concentrated on the density function. In Figure 7 can be seen the bivariate normal distribution function. The height of this figure represents the probability that a pair of $Z_1, Z_2$ values is below the $Z_1$ axis and below the $Z_2$ axis. Figure 7 shows the case where the correlation is .95. The ridge is less pronounced with a lower correlation. This function was generated by calling an IMSL subroutine that evaluates double integral areas. The subroutine was called 2601 times, once for each unique $Z_1, Z_2$ combination. The computer time for getting the height of the figure was $7150!$

Next we come to a distribution familiar to many students of statistics, the Student's $t$ distribution. The $t$ is a small sample, exact distribution, where it is assumed that the standard deviation and mean must be calculated from a finite population. As the size of this population approaches infinity, the distribution begins to resemble the normal. In Figure 8 is pictured the value of $t$ plotted against the number of degrees of freedom. As the number of degrees of freedom gets larger, it can be seen that the tails of the function contain less of the curve. Figure 8 is the density function for $t$, while Figure 9 is the distribution function. Here, it can be seen that to achieve a particular tail probability, say .05, a smaller value of $t$ is required as degrees of freedom get larger.

The difference between the density functions for the normal and $t$ distributions is pictured in Figure 10. The height approaches zero as the degrees of freedom get large. At a single degree of freedom, the differences between the two distributions are exaggerated. In particular, note that the $t$ has more in the tail than does the normal. The difference in the distribution functions is given in Figure 11.

A distribution related to the normal is chi-squared. The chi-squared is the sum of a normal squared. The chi-squared density function is shown in Figure 12. It is apparent that it begins to look more nearly normal at ten degrees of freedom. The distribution function is shown in Figure 13.

The $F$ distribution consists of the ratio of two chi-square variates. In Figure 14, we have held the numerator degrees of freedom constant, while varying the value of $F$ along one axis and the number of denominator degrees of freedom along the other axis. The height represents the density of the $F$. The average $F$ is unity, which can be seen in the figure. Figure 15 shows the distribution function for the $F$. The $F$ test is an exact test, while the chi-square test is made with an appeal to asymptotic properties where the sample is infinite. The relationship between the two is this: an $F$ with $n$ numerator degrees of freedom and $d$ denominator degrees of freedom will approach the value of chi-square with $n$ degrees of freedom divided by $d$ as $d$ approaches infinity. In practice, the two converge fairly rapidly, as can be seen in Figure 16. In the figure, the value of the numerator degrees of freedom is set to four. This figure is of the distribution function.

A function of particular mathematical simplicity, which is often used in place of the normal distribution, is called the logistic function. The logistic function can be seen in Figure 17. As in Figure 3, the variance is changing along one axis, the value of the function varies along the other, and the height of the volume is the density. This function is similar in shape to the normal except that the tails of the distribution are longer. In some situations it would be more realistic to fit the logistic rather than the normal. The logistic is often used with the analysis of proportions and growth data.

Figure 18 graphs the log-normal distribution. The log normal distribution is characterized by the log of the variable being distributed normally. In this figure, one of the axes represents the variable and the other represents the dispersion. The distribution is a normal distribution when sigma is zero, but diverges markedly from it as sigma increases. Since the function is a two parameter distribution one of the parameters must have its values held constant in a three-dimensional representation. The distribution has applications in the fields of economic and biological analysis. In addition, by proper selection of its parameters it can become a close approximation to the normal distribution, with the added property that the density can be made zero for all negative values of the variable.

The inverse Gaussian distribution is exhibited in Figure 19. This distribution arises from the sequential analysis techniques of A. Wald, and is sometimes known as the Wald distribution. It is also employed in the study of diffusion processes. A limiting form of this function is the normal distribution. In addition to its theoretical interest it was selected because of its esthetic properties.

Some distribution functions are so general that they can generate other distributions as special cases, i.e. by fixing certain values of their parameters. Such is the case of the two parameter gamma distribution, defined with parameters alpha and beta. Figure 20 is the exponential distribution and is equivalent to the gamma distribution with alpha equal to 1. Setting beta equal to 2 will form the chi-square distribution where the degrees of freedom are two times alpha. The normal distribution can be approximated by the gamma distribution if sufficiently large values of alpha are used. Figure 21 displays the form of this distribution in three dimensions while Figure 22 gives a two-dimensional picture. The gamma distribution is found in the analysis of random processes in time and life expectancy experiments.

The Weibull distribution is shown in Figure 23. If a power transformation is applied to the observed variable, the transformed variable is distributed according to the exponential
distribution. The distribution allows for the relaxation of the assumption of independence. It is used in quality control experiments and in the analysis of discrete data.

Also related to the exponential distribution is the Laplace distribution of Figure 24. As can be seen it is a double, or folded, exponential where one axis represents the variable and the other contains the standard deviation. Again this was selected as much for its appearance as for any theoretical interest.

The two parameter beta distribution can be used to approximate a large number of other distribution functions, including the F distribution. Along with the gamma distribution this function can generate distributions with a large number of different shapes. (See Figures 24 through 26.) Currently the beta distribution is receiving much attention in Bayesian analysis as a prior distribution for a binomial process. Figure 27 shows a two-dimensional representation with overlaid plots of the beta distribution for different values of the parameters alpha and beta.

REFERENCES


Figure 13

Chi Square

Figure 14

The $F(4,DF)$ Distribution

Figure 15

The $F(4,DF)$ Distribution

Figure 16

The difference between $F(4,DF)$ and Chi Square(4)/4

Figure 17

LOGISTIC DISTRIBUTION

Figure 18

LOGNORMAL DISTRIBUTION $\theta=.55$