

Statistical Distributions: How Deviant Can They Be?

James B. McDonald

There are many jokes told about economists and statisticians. A neighbor took great delight in telling me that he had heard that an economist is someone who wanted to be an accountant but didn't have enough personality. You may have heard about the statistician who drowned in a river that only averaged six inches deep.

In spite of the feelings these expressions reflect for these two disciplines, the use of statistics provides an important common denominator for much of the applied research being done in the natural and social sciences. This is true in the fields of physics, astronomy, biology, economics, engineering, finance, marketing, and many more. Statistics often refers to both the collection of empirical data and the use of this data to estimate relationships, determine trends, and make inferences. It is the analysis of data, however, rather than the collection of data that characterizes modern statistics.

Given the extensive data sources currently available and existing computer hardware and software, the use of statistical models to describe the extent of our uncertainty about a variable and relationships between variables is a particularly exciting and productive area of research. I will focus on one facet of my research which I hope will be of rather general interest. First, I will review some history associated with the development of statistics and probability from being purely descriptive to providing models for the analysis of data. Second, I will discuss a few of the statistical models known as distributions that have played an important role in the development of modern statistics. Major scientific problems have provided the basis for the development of several of these. Next, I will discuss some relatively new distributions that include most of the previous ones as special cases and also provide for important increased flexibility. Finally, I will consider some applications of these models in economics, engineering, and finance.

James B. McDonald is a professor of economics and managerial economics at Brigham Young University. This essay was originally presented as the twenty-fourth annual Distinguished Faculty Lecture at Brigham Young University, 28 January 1987. The author expresses appreciation to Steve White and Dave Williams for their able research assistance and to Steve for preparing many of the figures used in the presentation; to Earl Faulkner for providing references to some excellent material on the history of statistics; to Jay Irvine who provided data on starting salaries for graduates; to Richard Butler, Kaye Hanson, and Steve White for their comments on earlier versions of the paper; and to his parents, Leonard and Arola McDonald, and his wife, Kathy McDonald, for their assistance and encouragement.

THE HISTORY OF PROBABILITY AND STATISTICS

Attempts to determine the beginning of any intellectual discipline are highly speculative. Many researchers suggest that before 1650 statistics primarily involved the description of events rather than reasoning from the data in order to make inferences about cause and effect.¹ Probability theory is no different. There is evidence of games of chance being played long before the seventeenth century. Many archaeological finds contain abnormally large deposits of small ankle bones. It is thought that these bones were used in various games. These bones appear in paintings on Egyptian tombs as pieces used in games of chance and are still used in some children's games in France and Greece.² Gambling or gaming was so popular with the Romans that laws were passed which forbade it except during particular seasons. The emperor Claudius was so interested in "dicing" that he wrote a book about it and played while riding in his carriage. It is reported that he would throw dice onto a special board that had been fitted in his carriage. One source reports that he even played his left hand against his right hand.³

There appears to have been little or no formal discussion about the odds or probabilities associated with games of chance prior to the mid to late seventeenth century.⁴ The sample mean or average is very important in statistics and probability and yet isn't even mentioned before the eighteenth century.⁵

Why was the theory of probability and statistics so late in being developed? One explanation is that the dice or instruments used in gambling were so irregular in shape that it may have been difficult to recognize a consistent pattern from one set to another. The faces of the dice were often neither square nor parallel. F. N. David obtained three dice from the British Museum. One was made out of rock crystal, another from iron, and the third from marble. The three dice were each tossed 204 times, and the number of times that each of the six faces appeared was recorded.⁶ These results are shown in table 1.

TABLE 1
Dice: Result of 204 Tosses

MATERIAL	VALUE OF TOSS					
	1	2	3	4	5	6
Rock Crystal	30	38	31	34	34	37
Iron	35	39	30	21	37	42
Marble	27	28	23	47	25	54
Expected Number	34	34	34	34	34	34

Looking at the last row in table 1, we see that the expected frequencies associated with tossing a "fair die" 204 times would be for each side to appear thirty-four times. It should be apparent from the other rows that

there are obvious discrepancies between the observed and expected frequencies. Those differences are statistically significant for the marble die.

Another explanation for the relatively late development of a theory of probability and statistics is that until relatively recent times events in the world were viewed as being random or predetermined. The Greeks and Romans viewed the world as being partly determined by chance, with the gods and goddesses having some control over the outcome of events. A contrasting view is often attributed to early Christianity before the Reformation. The notion of a deterministic world without randomness or chance appears to have been quite common.⁷ Both of these views of the occurrence of events in the world would discourage the careful analysis of random events that is at the very core of modern statistics.⁸

It is interesting to note that the Bible, Book of Mormon, and Doctrine and Covenants contain references to the use of drawing lots as a method of making decisions as well as providing a way for the expression of God's will. The Jewish Talmud also includes many references to the use of lots. One of the more interesting is the description of the division of Israel among the twelve tribes. As reported by Hasofer, the procedure was as follows: Eleazar wore the Urim and Thummim, while Joshua and all of Israel stood before him. An urn containing the names of the twelve tribes, and an urn containing descriptions of the boundaries were placed before him. Animated by the Holy Spirit, he announced the name of a tribe and the name of a territory. Then he shook the urns and drew out the name of the tribe from one and the territorial description from the other. This procedure was repeated for each of the tribes. It is interesting that the results of the drawing were reported to have been announced prior to the drawing and that at least one source reports that the drawings from the two urns involved two priests. These elaborate preparations were to emphasize that the results were the outcome of divine will.⁹

Two other excellent references to the history and development of probability would be the books by Hacking and Maistrov.¹⁰

WHERE DO OUR MODERN STATISTICAL MODELS COME FROM AND HOW ARE THEY USED?

The development of formal statistical models began in the late seventeenth century. Some of the early developments were motivated by problems in astronomy and physics. The participants in this development represented many disciplines, and a list of the contributors reads like a *Who's Who in the World of Science*, including Euler, Edgeworth, Demoivre, Galton, Gauss, Laplace, Legendre, and Maxwell, among others. These contributions arose out of an attempt to build models that

would more accurately describe some phenomenon and use the available data more efficiently to make decisions in the face of uncertainty. The notion of probability or distribution functions provides the theoretical bedrock or foundation for these efforts.

The distribution or density function is used to visually depict the relative frequency, likelihood, or probability that an event will occur. Some important concepts are humorously depicted in figure 1, drawn by Ron Bell. In this figure, those above 90 percent receive A's, those between 80 and 90 percent receive B's, and so on. The corresponding areas under the curve indicate the fraction of students receiving the various grades. This particular distribution is symmetric with both tails having the same shape and thickness.

The shape and location of these distribution functions is very important in statistics. For example, in 1986 the average starting salary was almost \$28,000 for a student with a bachelor's degree in engineering and \$16,000 for those graduating with a bachelor's in social and recreation work. Figure 2 visually illustrates that not everyone received the same salary; however, there is a \$12,000 difference between the average starting salaries for the two majors. If Steve Young had majored in recreation work and graduated in a class of fifty, the average salary for that major would be approximately \$36,000, as indicated in figure 3. If the graduating class was smaller and only included eight students, then the average starting salary would be approximately \$136,000 per year. Figure 4 illustrates a distribution with a thick tail to the right and is said to be *skewed* to the right. Starting salaries for MBA's exhibit this same behavior, with a few students offered in excess of \$70,000 per year.

Figure 4 provides a different example. The grade inflation issue at BYU can be viewed as a distribution that has a thick tail to the left and is said to be skewed to the left. This can be due either to professors who are too lenient or to a relatively large group of excellent students.

The shape of the distribution, whether it is skewed or not, and the thickness of the tails have very important consequences when we attempt to model uncertainty. For example, many students seem to feel that their entire futures depend upon the shape of the curve used in determining final grades. The thickness of the tail indicates the probability of large deviations from the mean. One can imagine that this would not only be of interest to students, but to a portfolio manager who is interested in large returns (right tail) but is also concerned about the likelihood of large losses (left tail).

Important statistical distributions developed before the twentieth century include the uniform, binomial, normal, beta, double exponential or Laplace, chi-square, lognormal, Student's t, and Pearson's skew distributions. I will not be exhaustive in my coverage of these distributions. However, some fascinating stories are behind the development of

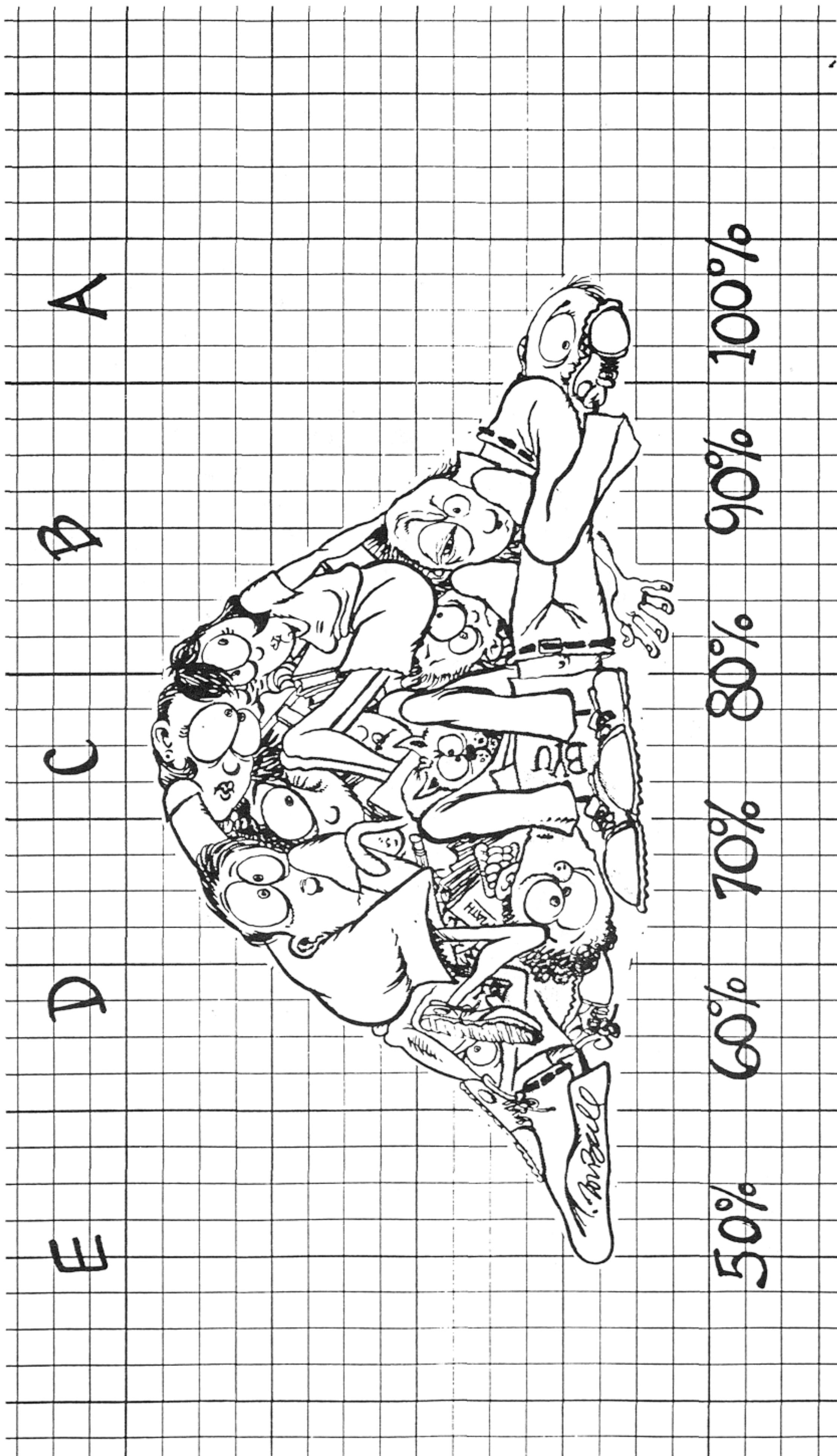


FIGURE 1

FIGURE 2

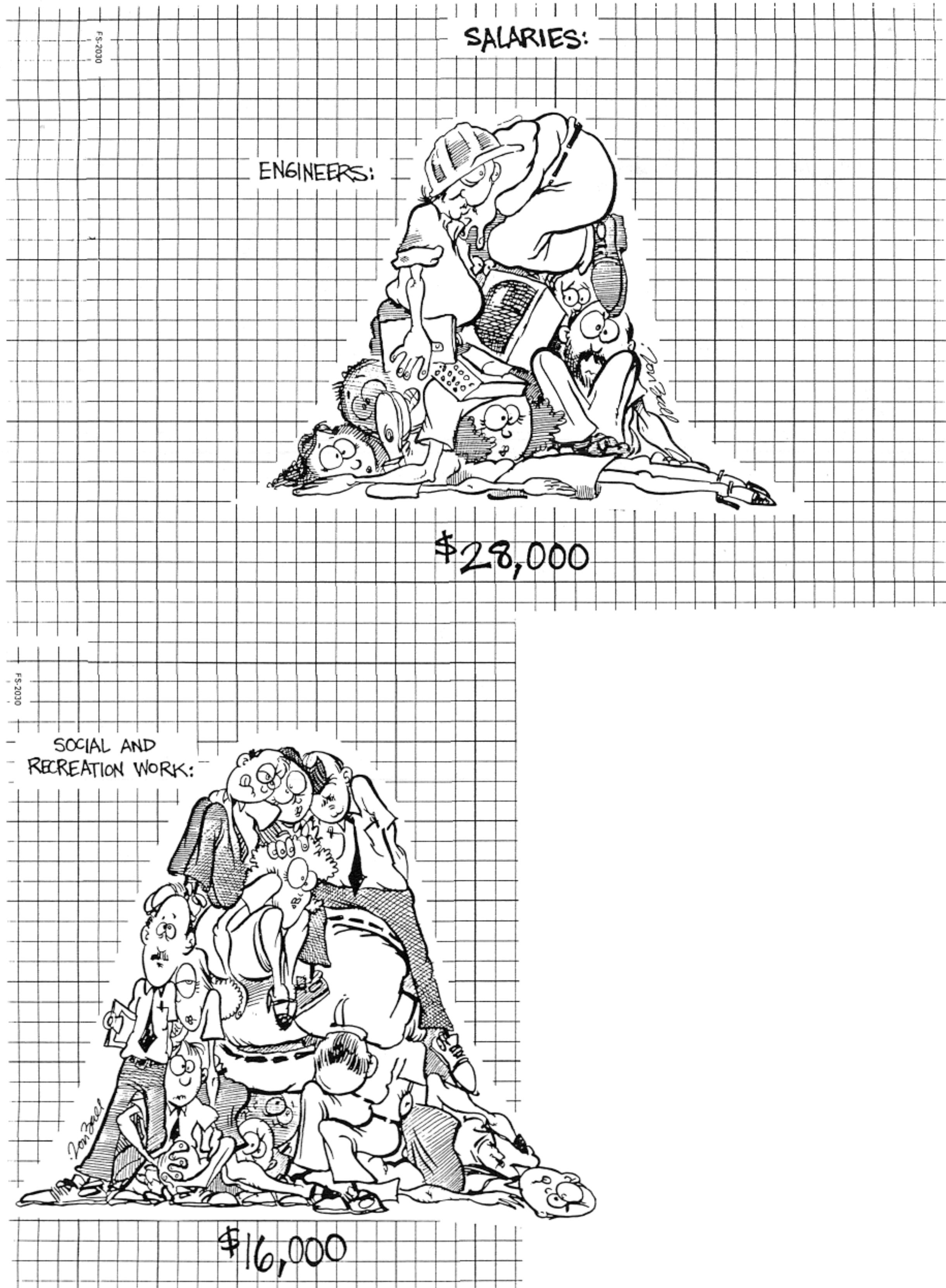


FIGURE 3



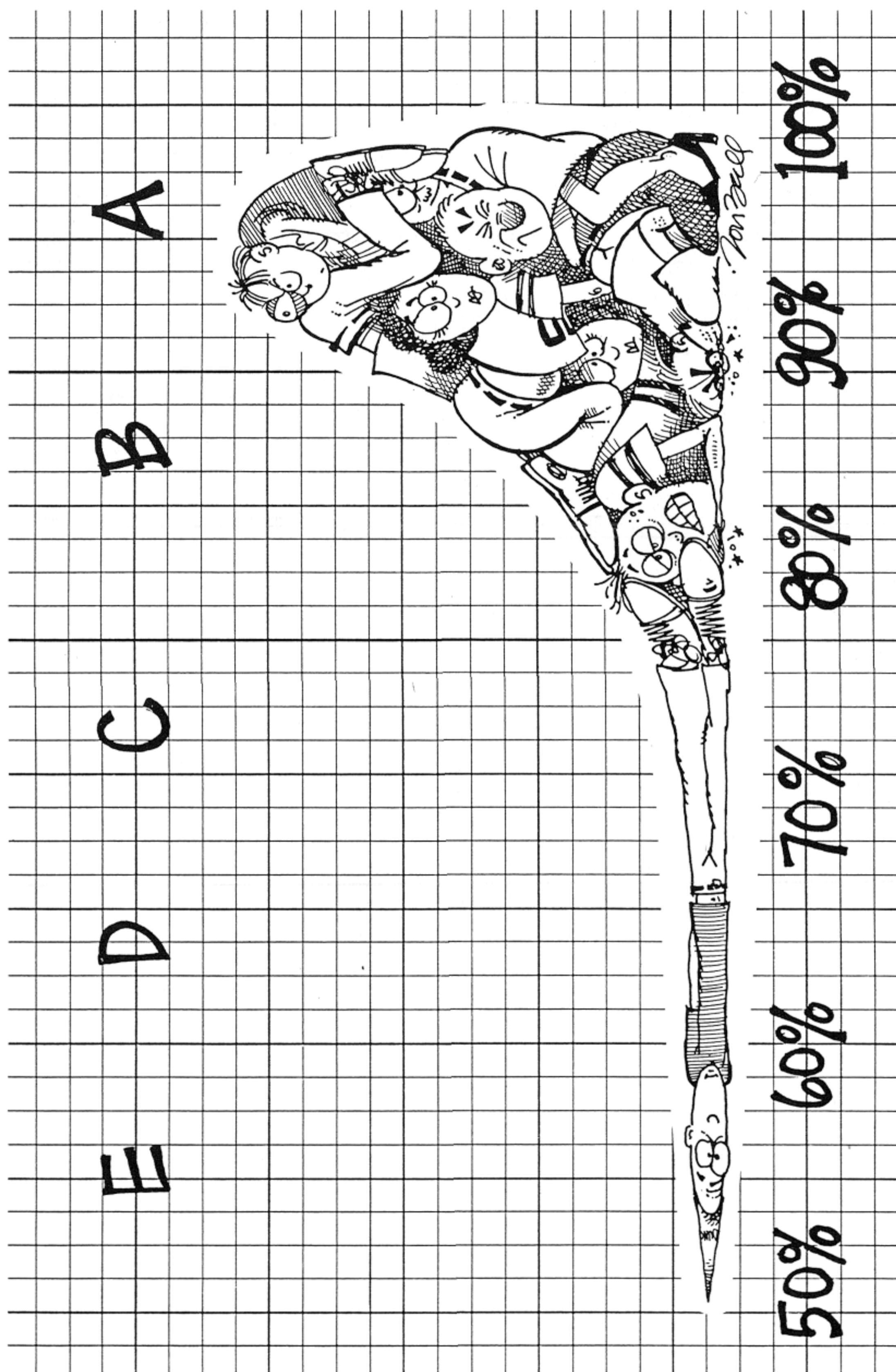
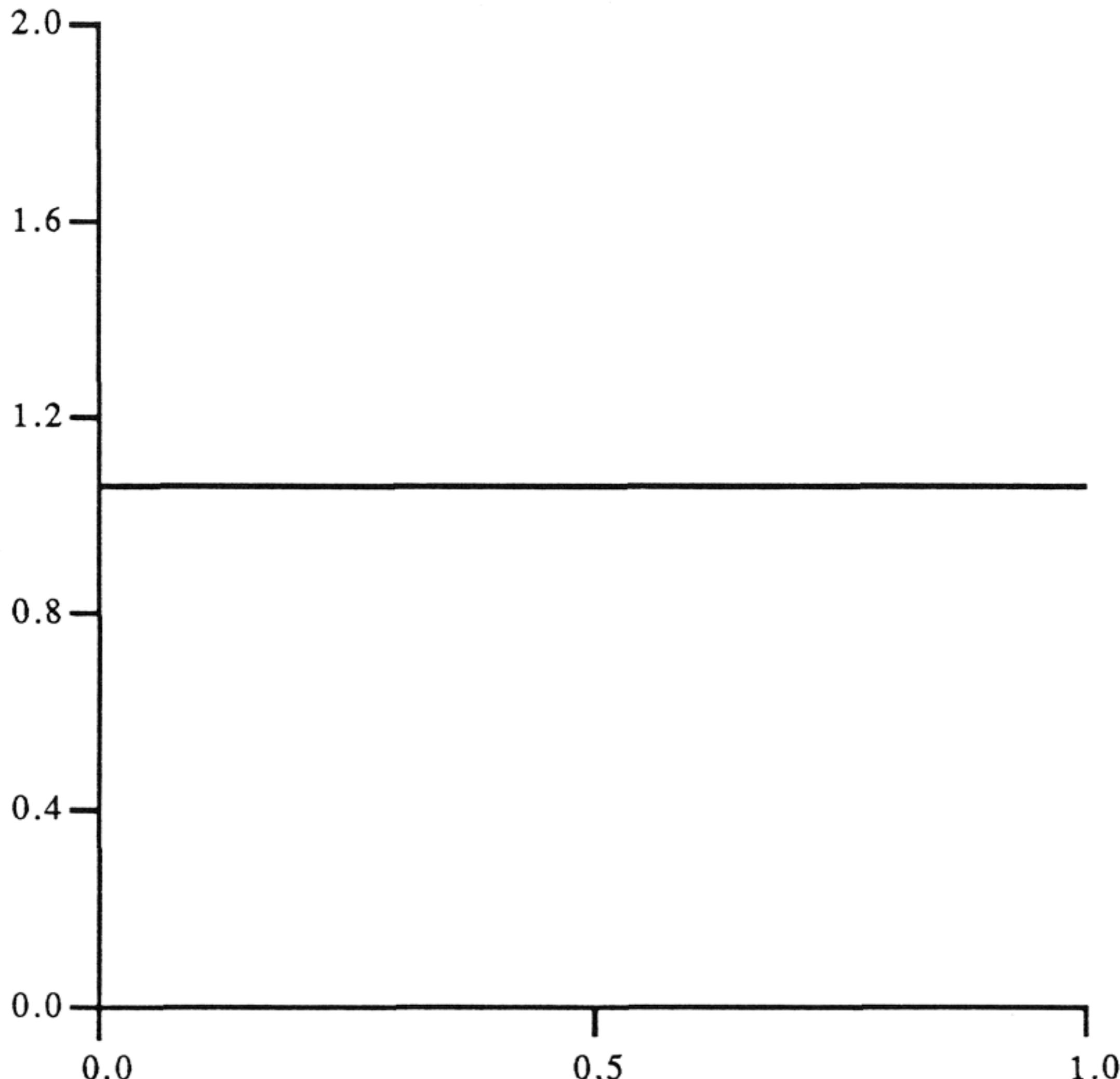


FIGURE 4

some of these models. I will briefly trace the evolution of a few statistical distributions and focus on their shapes. I will not address normative issues associated with the shapes of the distributions that arise in various applications.

FIGURE 5
Uniform

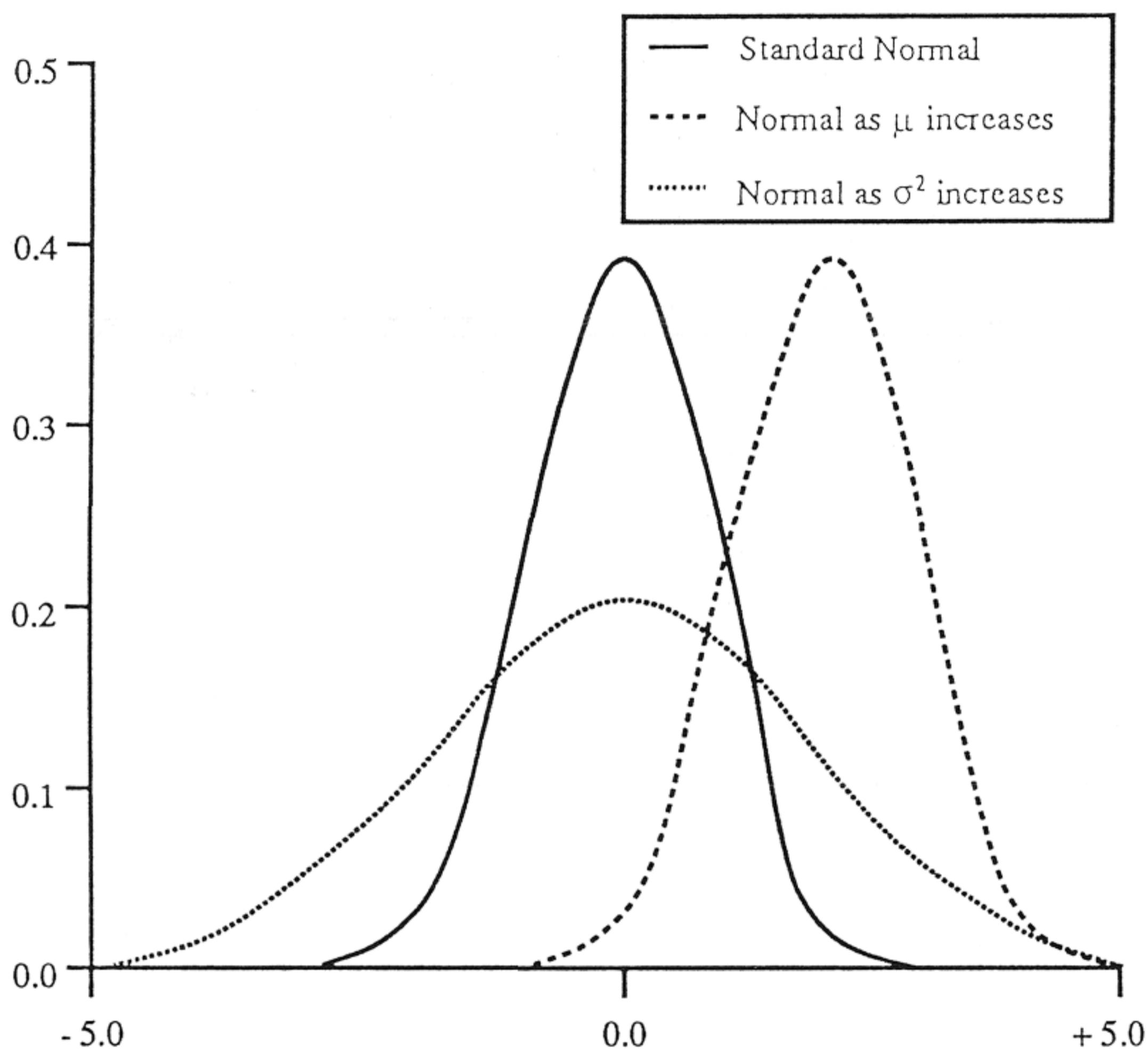


One of the first statistical distributions to be observed and mathematically modeled is the uniform distribution shown in figure 5. This distribution appeared as an estimate of an empirical law in some of Halley's data on human mortality.¹¹ DeMoivre formalized the distribution in his treatise on life annuities.¹² The model suggests equally likely outcomes of an event over a finite interval, and the average or expected value of the event is at the midpoint of the interval.

The famous bell shaped curve, the normal probability distribution with various possible shapes, is shown in figure 6. Distributions with different means (μ) and variances (σ^2) are shown. In 1733, DeMoivre first obtained the normal distribution as an approximation to binomial

distribution.¹³ Only later was it found to provide an excellent fit to many types of data.¹⁴ This density was often called “the law of frequency of error” and is one of the most commonly used distributions in statistics—the workhorse of statistics. The mean or average value of a normally distributed variable corresponds to the highest point on the curve, and it is important to note that the distribution is symmetric about the mean.

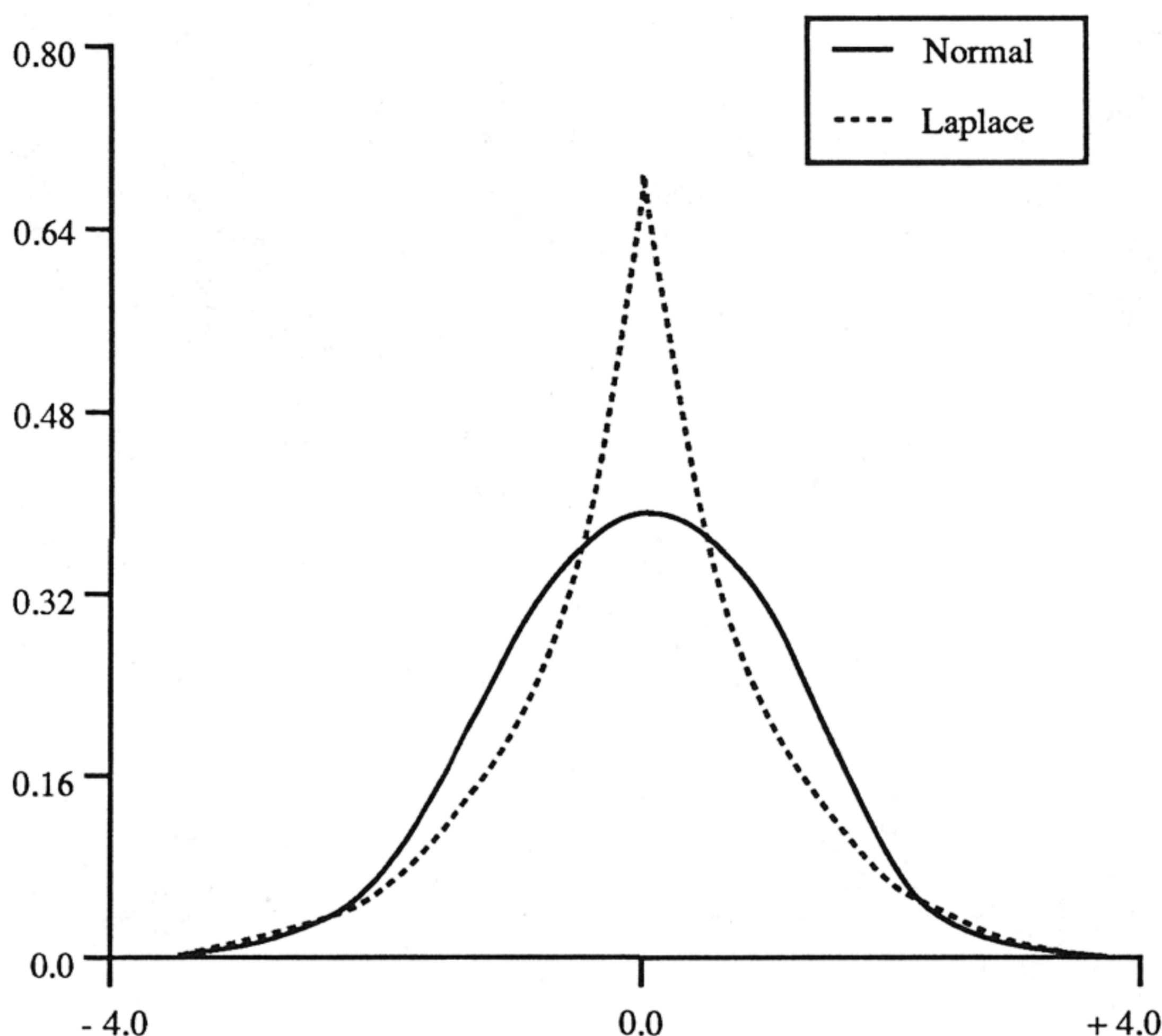
FIGURE 6
Normal



It is difficult to assess the impact the normal distribution has had on theoretical and applied statistics. However, the potential of the normal distribution was recognized early. In 1889 Francis Galton wrote in his famous book *Natural Inheritance*: “I have known of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the Law of Frequency of Error (normal). . . . The Law would have been personified by the Greeks and deified, if they had known of it.”¹⁵ Historically, the development of the normal distribution or theory of errors was particularly associated with astronomy, but it is now widely used in many disciplines.¹⁶

However, the normal has two important shortcomings: many data are not symmetrically distributed, and some distributions have a higher frequency of outliers or thicker tails than permitted by the normal. In fact it seems that skewed distributions are often the rule, rather than the exception for many economic data. Two distributions that permit thicker tails than the normal are the Laplace and Student's t distributions.

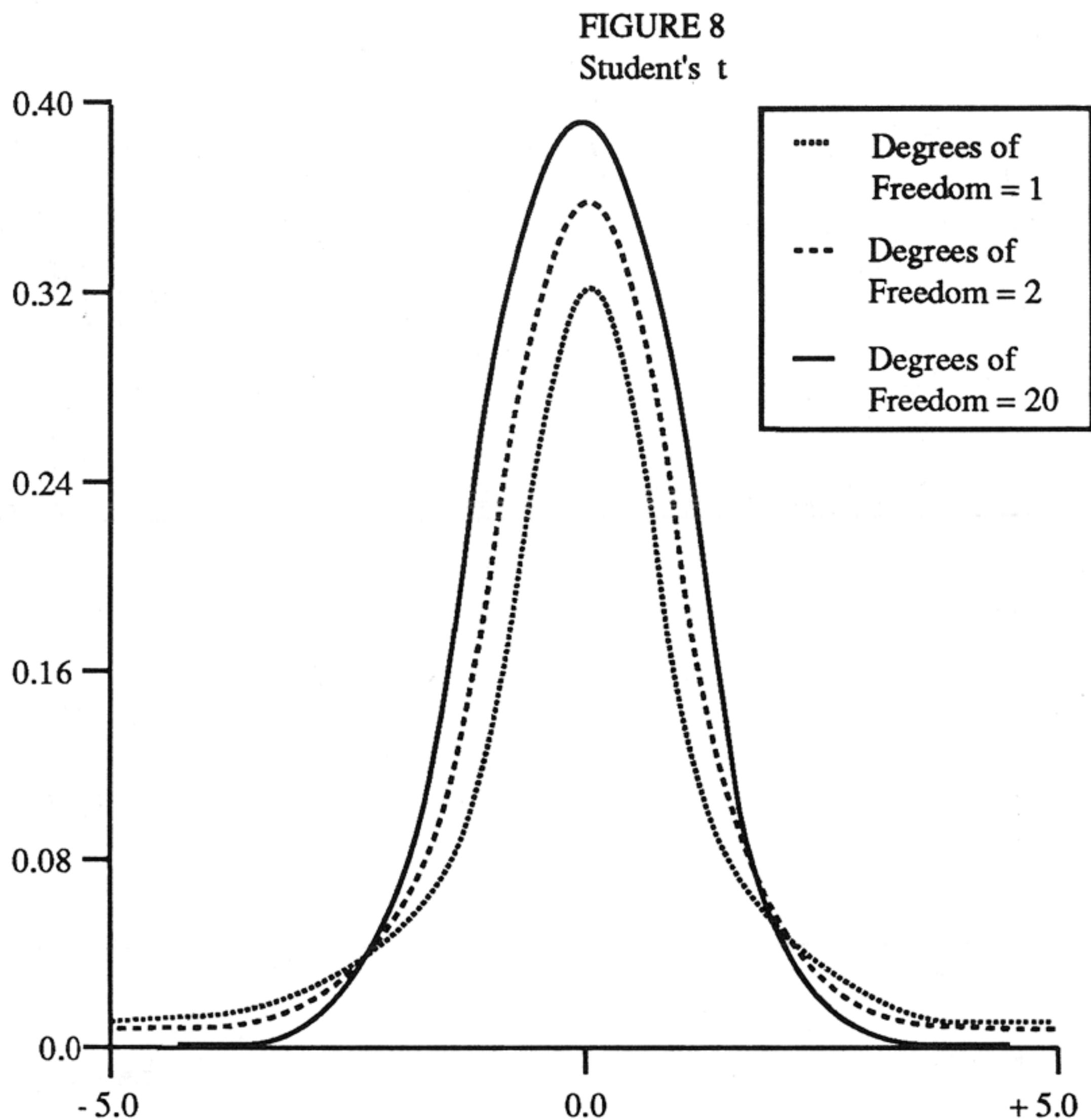
FIGURE 7
Normal and Laplace



In 1774, Laplace derived the double exponential or Laplace distribution. From figure 7 we see that the Laplace distribution, like the normal, is symmetric about the mean but is more peaked near the mean and has thicker tails. Both of the distributions in figure 7 have the same variance. An interesting personal note about Laplace is that he was one of Napoleon Bonaparte's instructors and had interesting public and scientific careers. He served as minister of the interior in 1799, and later as a member of the French Senate. Laplace was a productive scholar until his death at the age of seventy-eight and has been referred to as France's most illustrious scientist of the eighteenth century. He was eulogized by

Poisson as “the Newton of France.”¹⁷ One biography suggested that if “publish or perish” were literally true, Laplace would be alive today.

The problem of thick tailed data led to the development of another statistical model that came from the Guinness Brewery in Ireland.

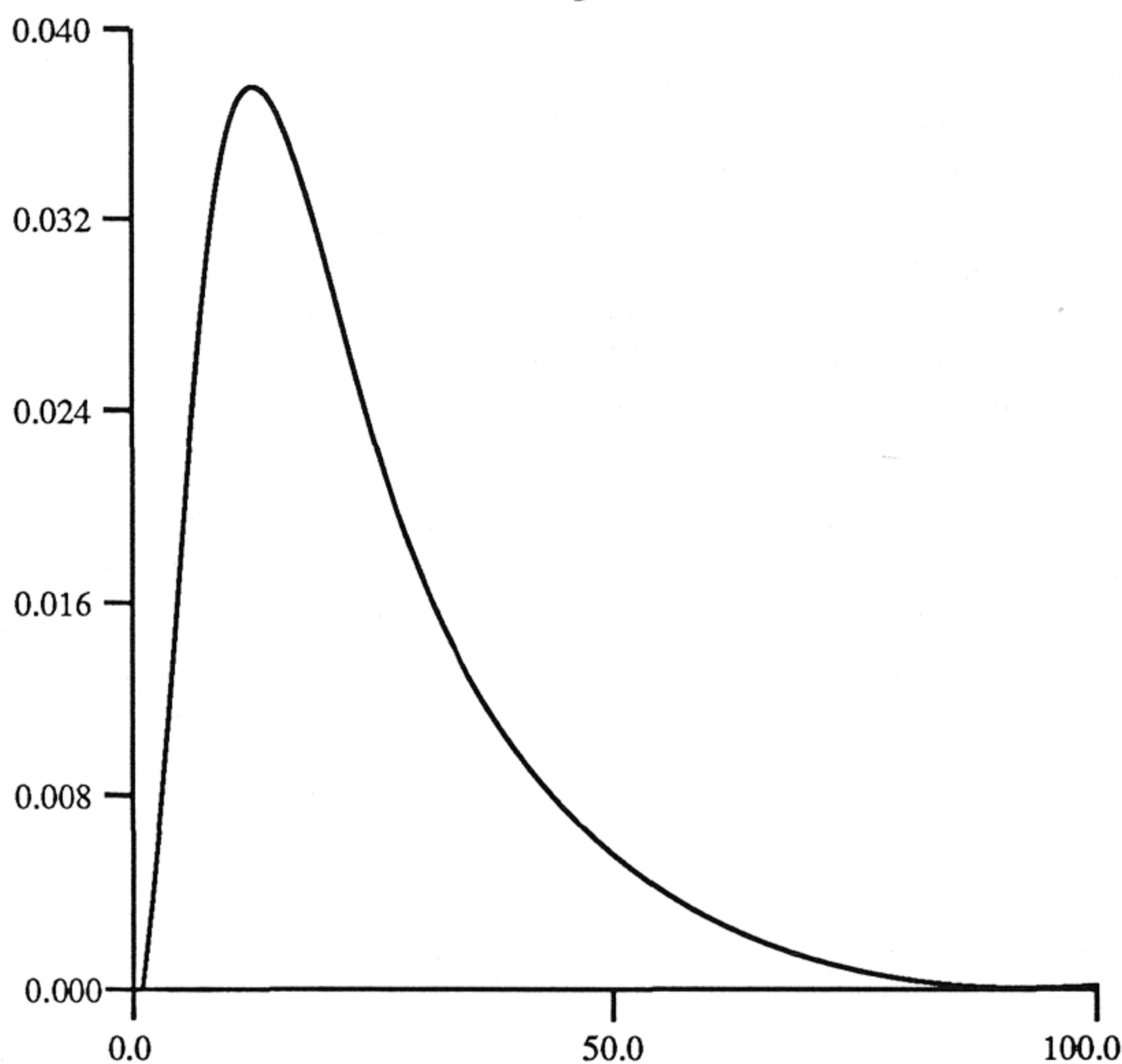


The t distribution is another distribution that is symmetric but permits thicker tails than the normal. The smaller the degrees of freedom, the thicker the tails. The normal distribution is a limiting case of the t distribution for large degrees of freedom. This distribution was derived by William Gosset. Gosset (1876–1937) graduated from Oxford in chemistry and mathematics, and was hired by the Guinness Brewery to study the production of beer and to investigate the relationship between the quality of the brew and the conditions of production. The normal distribution did not have thick enough tails to provide an accurate description. Gosset derived the famous t distribution and published his findings anonymously under the name “Student” in 1908. Since he didn’t

use his real name, he must not have been worried about tenure or promotion at the brewery. Gosset worked at the brewery until three years before his death, and his most important contributions in statistics were motivated by a desire to solve problems encountered at the brewery.

While these distributions helped with the problem of thick tails, neither provided a model for the positively skewed distributions that are so common in empirical work. The lognormal is a very important distribution for data that are skewed to the right.

FIGURE 9
Lognormal



In 1879, Francis Galton presented a paper before the Royal Statistical Society in which he stated: "My purpose is to show that an assumption which lies at the basis of the well-known law of Frequency of Error [the normal] is incorrect in many groups of vital and social phenomena."¹⁸ He then proposed the lognormal distribution for such data. If data are distributed as the lognormal distribution, then the natural logarithms of the data will be normally distributed. This

distribution is positively skewed with a long tail to the right and has been used extensively to model the distribution of income, particle size in engineering, and also in medicine. The lognormal can approximate the normal in some instances, but it cannot model negatively skewed data and often does not have thick enough tails.

In spite of the problems of asymmetric and thick-tailed data, the normal was often used rather uncritically until the early 1900s. Karl Pearson, among others, was very concerned about the shortcomings of the normal distribution and derived a system of distribution functions that permitted much greater flexibility than the normal. Pearson recorded in 1895 that Edgeworth had come to him about two years earlier with some skew price curves and asked if he could discover any way of handling skewness. Pearson reports: "I went to him in about a fortnight and said I think I have got a solution out, here is the equation, and told him my chief discoveries. I further said I don't intend to publish till I have illustrated every point from practical statistics."¹⁹ This system of distribution functions is still important today and includes the t-distribution, gamma, beta, and others as special cases.

In summary, the question of whether the normal fits the data well is important in many applications. There have been many attempts to address questions about normality, skewness, or symmetry, and the width of tails as measured by what is referred to as "kurtosis." The Laplace and t distributions provide some additional flexibility for the problem of the tails. The lognormal and Pearson skew distributions provide an approach for the skewness problem. Issues surrounding these developments have evolved over more than two hundred years and find roots in many disciplines.

The computational aspects of statistical analysis have been a major obstacle until fairly recent times. As an example of the time-consuming nature of complex calculations, the research for Karl Pearson's book *Tables of the Incomplete Beta Function* was begun in 1922, and the book was not published until ten years later in 1932. The dramatic changes in our ability to do complicated calculations that have occurred in recent years were unanticipated by many. For example, Charles H. Duell, of the U.S. Patent Office, is reported to have suggested in 1899 that "everything that can be invented has been invented" and even discussed closing the patent office. Thomas J. Watson, chairman of the board of IBM, declared in 1943, "I think there is a world market for about five computers." Popular Mechanics reported in March 1947: "Where a calculator on the ENIAC is equipped with 18,000 vacuum tubes and weighs 30 tons, computers in the future may have only 1000 tubes and perhaps weigh only 1.5 tons."²⁰

Today, a twenty pound portable computer has greater capacity than many of the large early computers that had to be kept in air-conditioned

environments because of the heat generated by the thousands of vacuum tubes and the sensitivity of the circuits to the physical environment. These recent developments in computer hardware and software have facilitated rapid progress in disciplines that are dependent upon numerous or complicated computations. Many new statistical models have been developed. The estimation and analysis of many of these models is often very complicated or impossible without the use of the computer. We now turn to some of these new distributions.

NEW DISTRIBUTION FUNCTIONS

Some of my recent research has dealt with very flexible distribution functions. If the wrong distribution function is selected, one can obtain very peculiar results. For example, if a normal distribution is fitted to highly skewed empirical data, important results can be in error and misleading.

The new distributions will be referred to as generalized beta of the first type (GB1), generalized beta of the second type (GB2) and generalized t (GT) distributions. The formulas for the distributions are given in table 2.

TABLE 2
Generalized Distributions

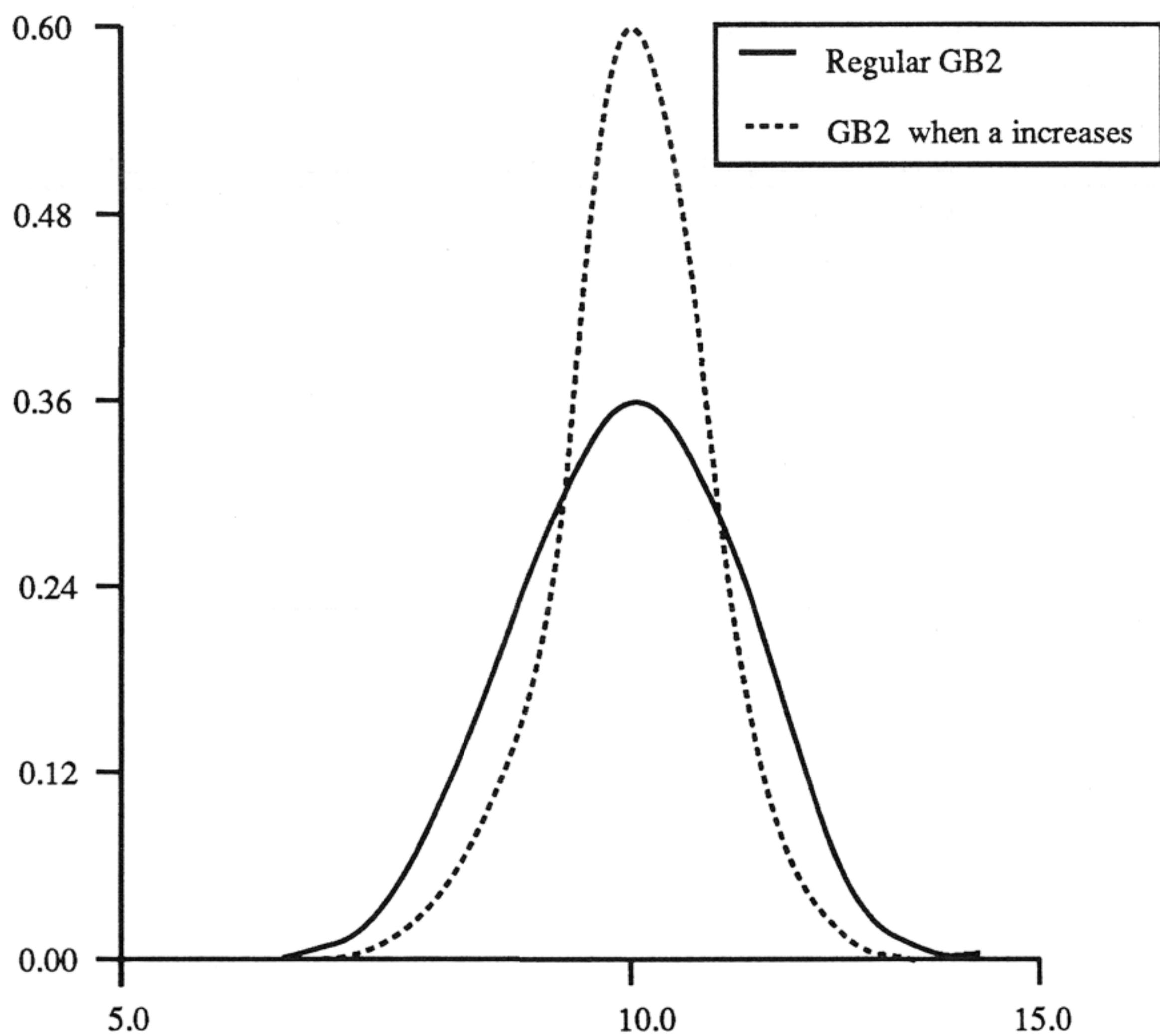
<u>MODEL</u>	<u>DENSITY FUNCTION</u>	
GB1	$\frac{ay^{ap-1}(1-(y/b)^a)^{q-1}}{b^{ap}B(p,q)}$	$0 < y < b$
GB2	$\frac{ay^{ap-1}}{b^{ap}B(p,q)(1+(y/b)^a)^{p+q}}$	$0 < y < \infty$
GT	$\frac{p}{2\beta q^{1/p}B(1/p,q)(1+(y /\beta)^p)^{q+1/p}}$	$-\infty < y < \infty$

These distributions are relatively "friendly" equations as compared to many equations in mathematics and statistics.

The GB2 or generalized beta of the second type has been considered before by Mathai and Saxena (1966) and Prentice (1975). However, this distribution was not widely known at that time, was independently obtained by several researchers in the early 1980s, and has received considerable attention during the last couple of years. The first applications of the GB2 to empirical data were done at BYU. The GB1 and GT do not appear in the literature and were developed at BYU and also have important applications in a number of different disciplines.²¹ The shapes of these distributions are extremely flexible and address many of the criticisms of the normal—in particular the issues of

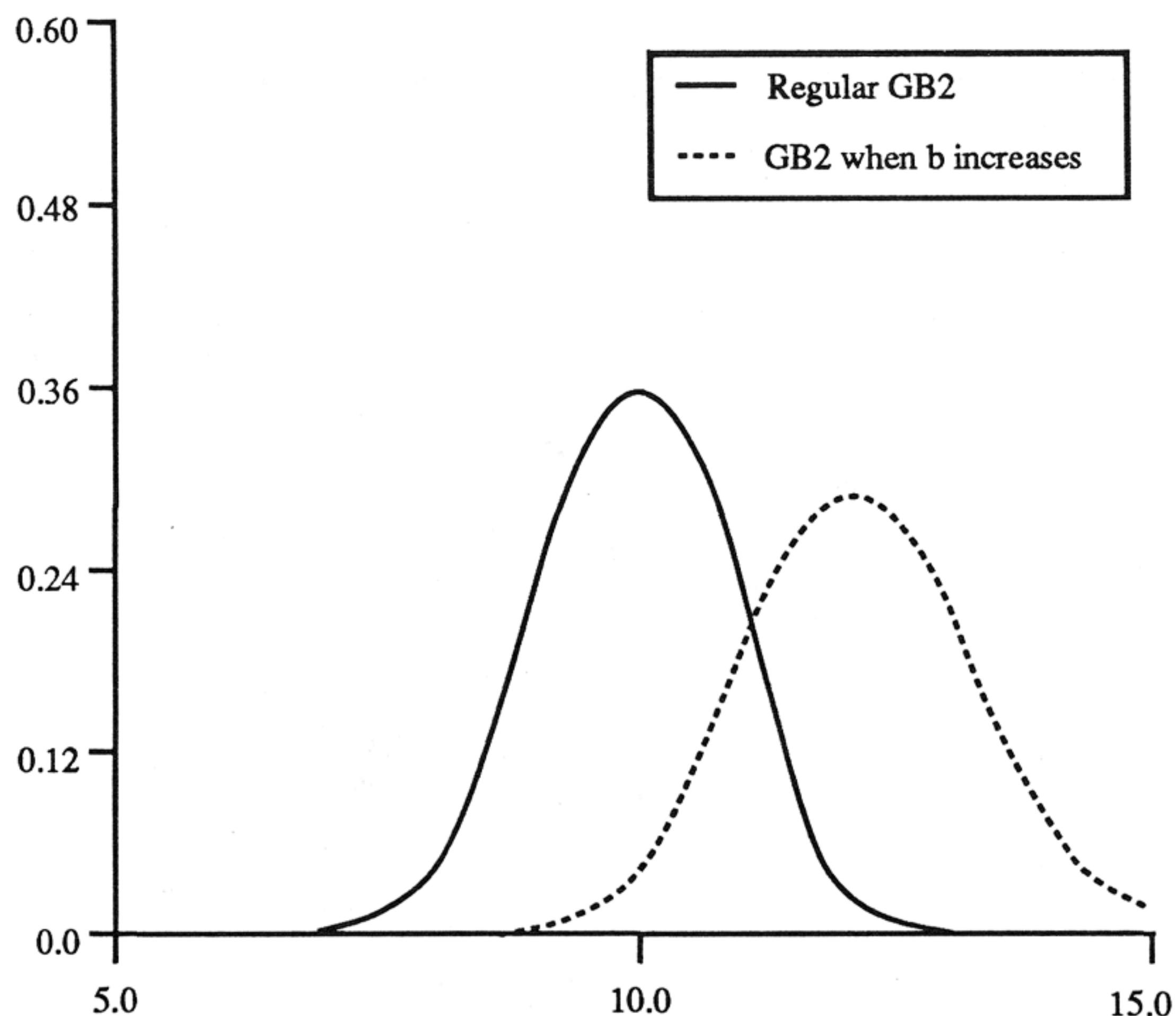
symmetry and thickness of the tails. For example, the GB2 includes four parameters (a, b, p, and q), and changes in these can accommodate four different types of movement of the distribution. These movements and flexibility are depicted in the following figures. Increasing the parameter “a” makes the distribution more peaked.

FIGURE 10
GB2 When a Increases



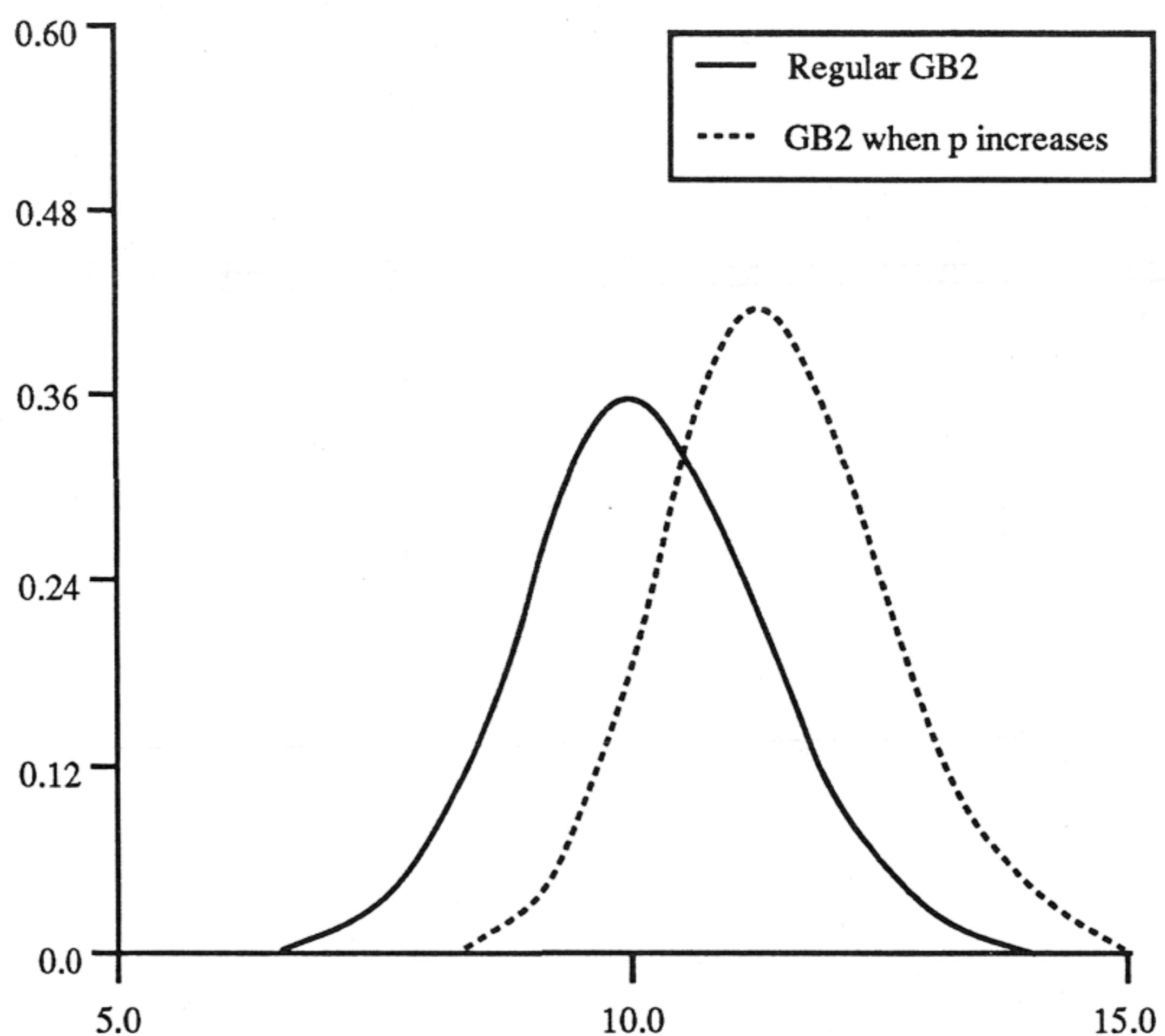
Increasing the parameter “b” shifts the distribution to the right.

FIGURE 11
GB2 When b Increases



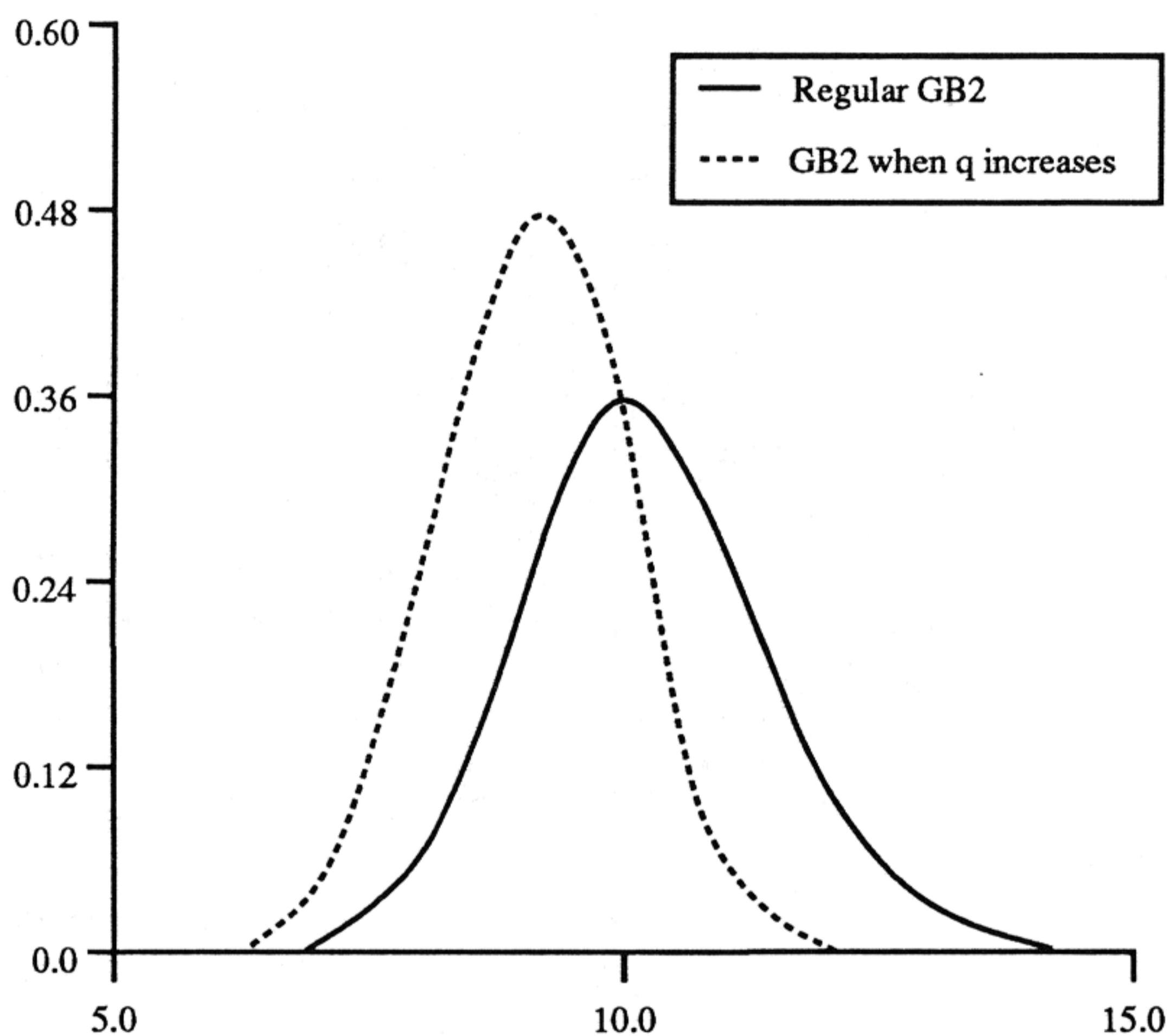
Increasing the parameter “p” tends to make the right tail thicker and the distribution more skewed to the right.

FIGURE 12
GB2 When p Increases



Increasing the value of "q" makes the left tail thicker and the distribution more skewed to the left.

FIGURE 13
GB2 When q Increases



In order to fit a distribution to a set of data, we use the computer to adjust the values of a , b , p and q and move the graph of the distribution until it fits the empirical data well. Adjustments of this kind would not have been feasible until recently, but computer programs have been developed to perform this estimation.

FIGURE 14
Distribution Trees

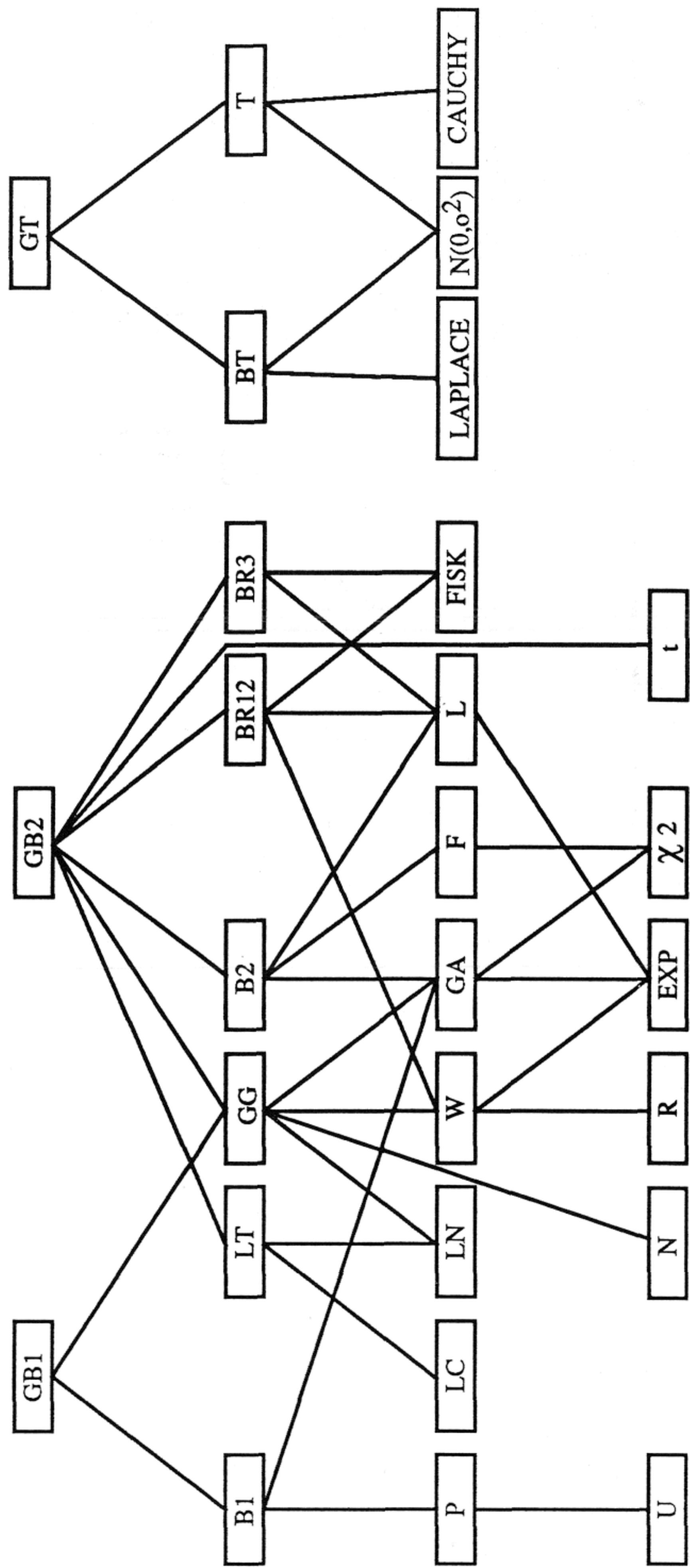
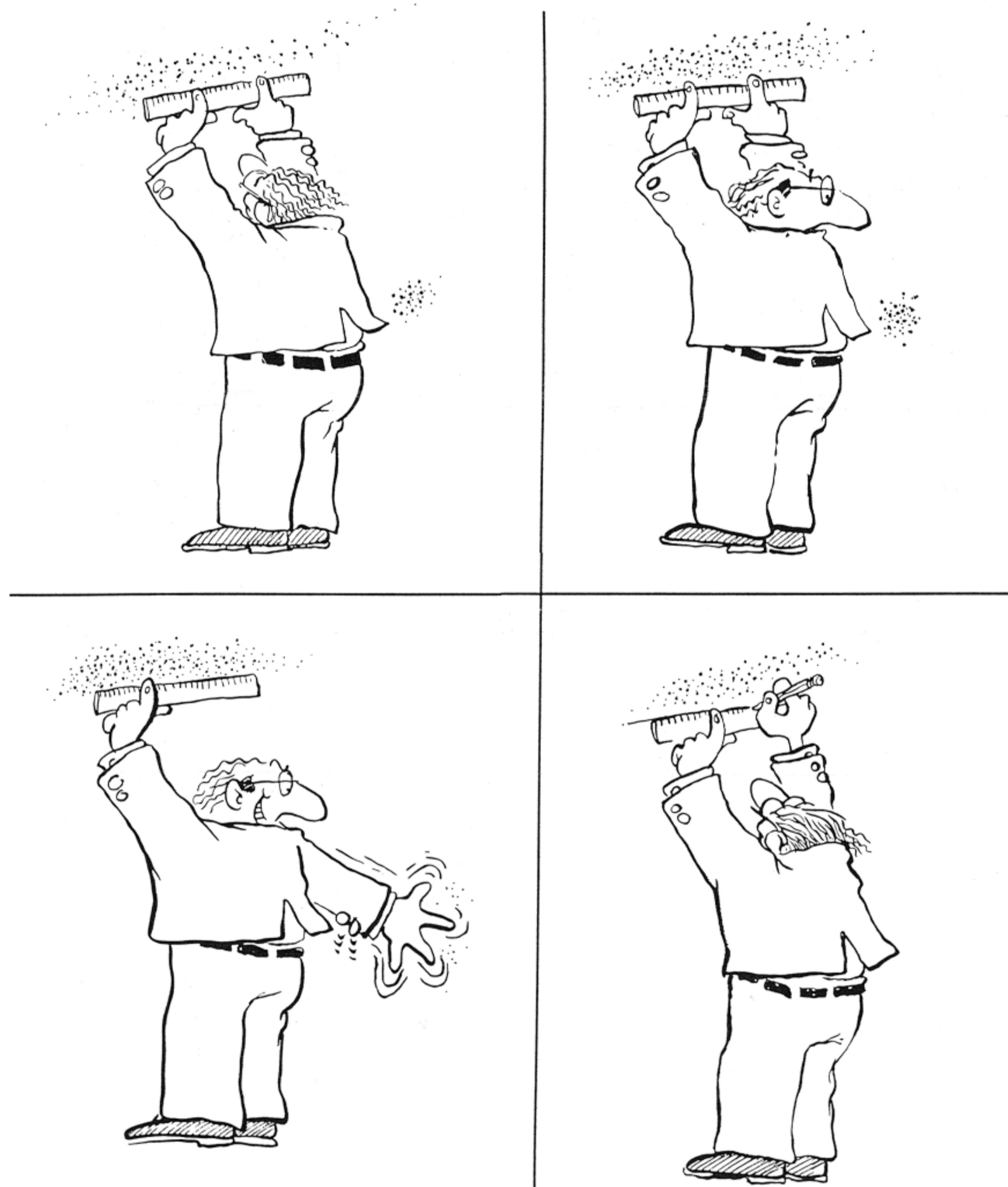


Figure 14 illustrates the flexibility of the general distributions in a different way. In this figure, each square or box corresponds to a different distribution. The connecting lines indicate special cases. The GB1, GB2, and GT can be seen to include many distributions as special cases. The GB1 and GB2 include the normal (N), lognormal (LN), gamma (GA), and Weibull (W) as special cases. The generalized t (GT) includes the Laplace, normal, t, and others as special cases. When I presented this material at a seminar at Princeton, someone asked if my Mormon background had motivated me to represent the relationships in the form of a genealogy tree.

The three general distributions include almost all of those used before as special cases and are extremely flexible in shape. By using the flexible or general distributions in empirical work, we can avoid imposing unrealistic assumptions associated with some of the special cases. The more general distributions will also fit any data set at least as well as any of the special cases. In the applications to be considered in the next section, I will focus on the GB2, but each of these three general distributions has important applications.

Before turning to the applications in the next section, I will briefly indicate how the generalized t can be used in regression analysis or curve fitting. Figure 15 depicts a common statistical problem that was involved in the solution of three important scientific problems in the eighteenth century: obtaining a mathematical model of the motion of the moon; determining the shape of the earth; and explaining the acceleration and deceleration of Saturn and Jupiter. With each of these problems, data were available that did not exactly conform to what was implied by the underlying models. The question can be visualized as trying to best fit a straight line to a set of data that do not lie on a straight line. One approach is to delete problem data points, as in figure 15, but even then the remaining data points will not lie on a straight line. Euler (1749) did not employ any statistical techniques. He simply noted that there was not an exact solution and moved on to consider different problems. Boscovich (1755, 1775) and Laplace (1785) proposed fitting a line that minimized the sum of absolute values of the vertical distances between the line and the observations. This procedure is based upon the Laplace distribution. Legendre (1805) proposed selecting the line that minimized the sum of squares of the vertical distances. This method is known today as least squares. Gauss (1809, 1823) proposed the same procedure and showed that this method is based upon the normal. The literature contains a rather lengthy and heated exchange between Gauss and Legendre as to who first discovered the method of least squares. Since the generalized t includes both the normal and Laplace distributions, it provides a generalization of both of these methods of estimation.²²

FIGURE 15
Regression Analysis



APPLICATIONS

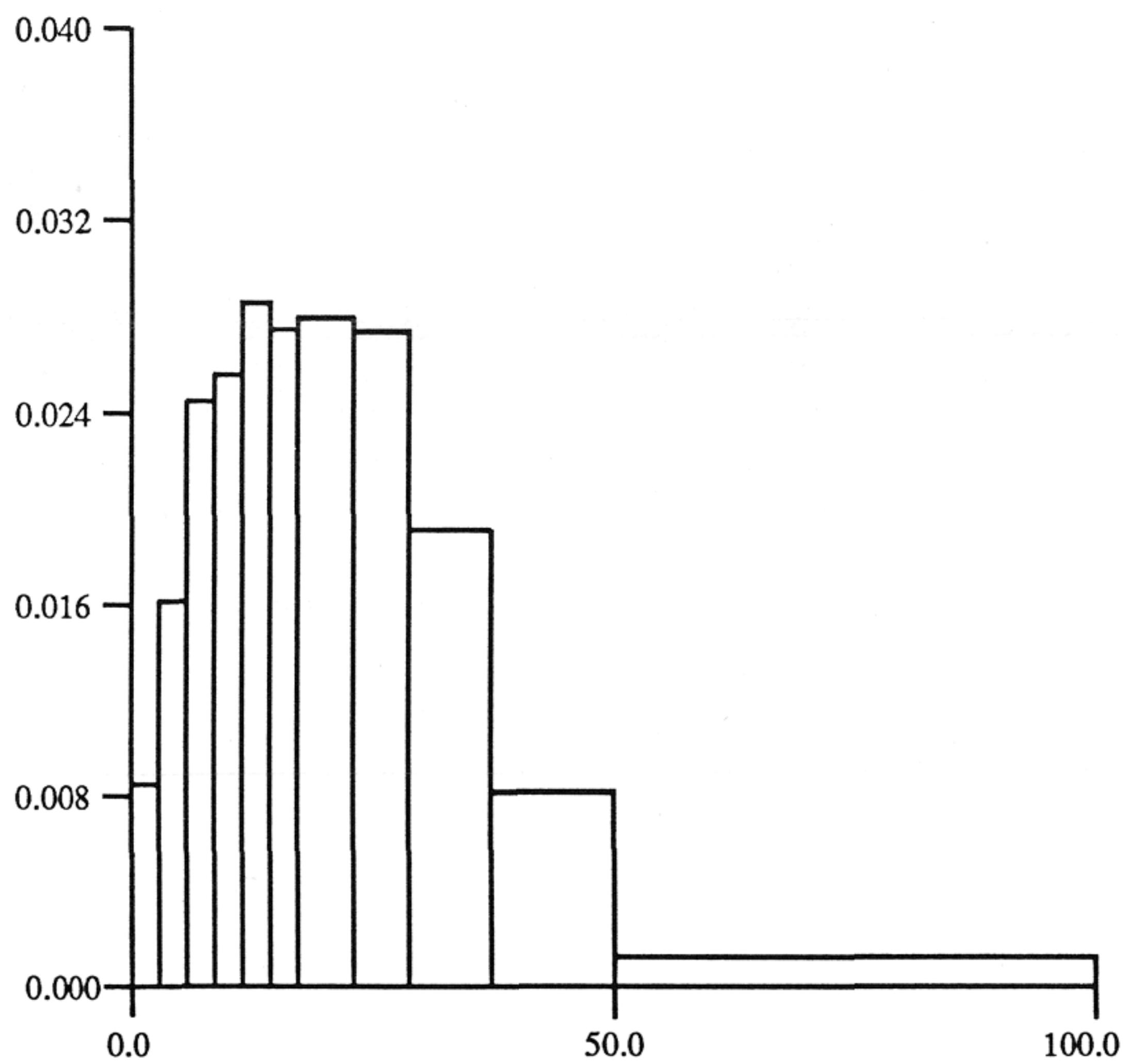
We will now consider three applications of the GB2: the distribution of family income, the distribution of coal particle size in coal combustion, and the distribution of stock prices. In each of these applications, we are not only interested in the mean or average value but also in the dispersion or variance, the skewness and thickness of the tails.

Income Distribution

Some of my early research involving statistical distributions was prompted by an interest in the distribution of income. Issues surrounding the distribution of income have attracted a great deal of attention from many economists and politicians, particularly since World War II. This is evidenced by considerable discussion of the impact of existing and potential economic policies upon various income classes. In order to provide answers to some of these questions, it is important to be able to quantify measures of income inequality in a useful manner and to investigate the relationship between these measures and important underlying macroeconomic and policy variables. It is also important to note that the distribution of income can have an impact upon the performance of the economy.

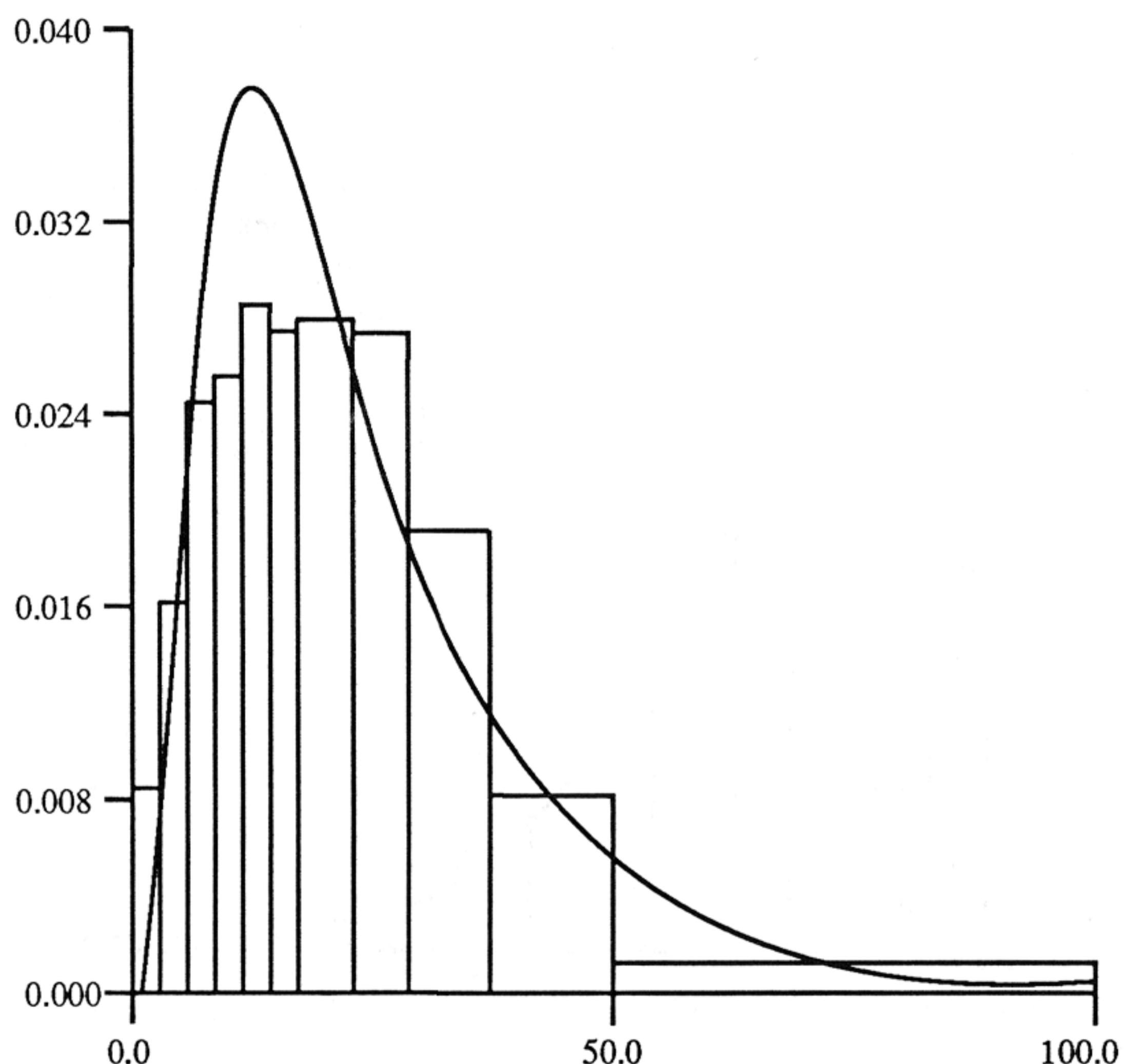
Many studies have considered these and related questions. These studies have often been based upon distributions that did not provide a good fit to the data or that used measures not sensitive to underlying changes in the distribution. For example, the data usually utilized in such studies are in a grouped format such as the distribution for family income for 1980 shown in figure 16. This is in the form of a bar graph with the areas of each "bar" representing the fraction of families in each income interval.

FIGURE 16
1980 Income



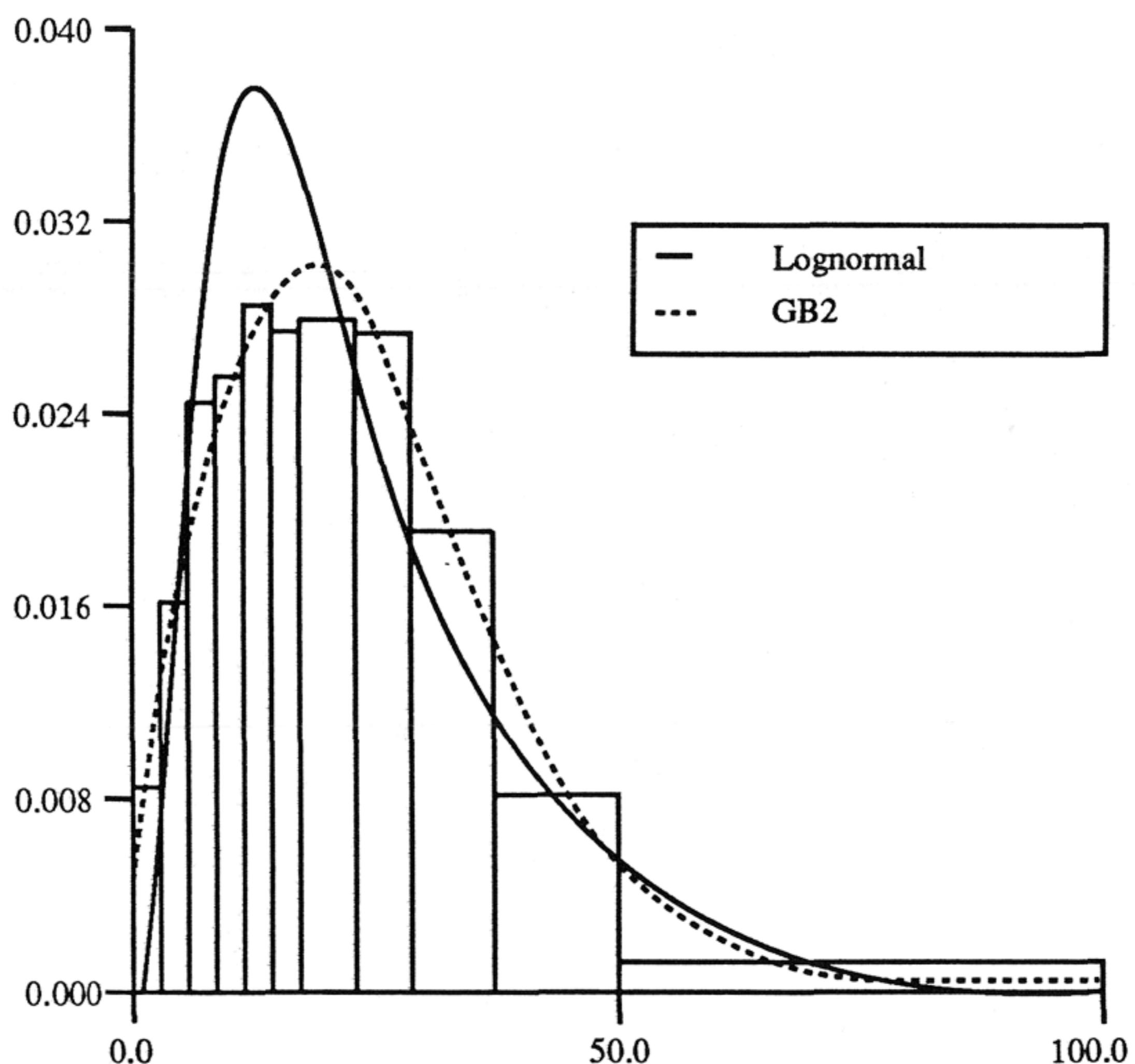
If a lognormal distribution is used, we obtain the fitted curve in figure 17. The lognormal fits some areas quite well, but not others. Note, for example, that the lognormal is too peaked near the center of the distribution.

FIGURE 17
1980 Income and Lognormal



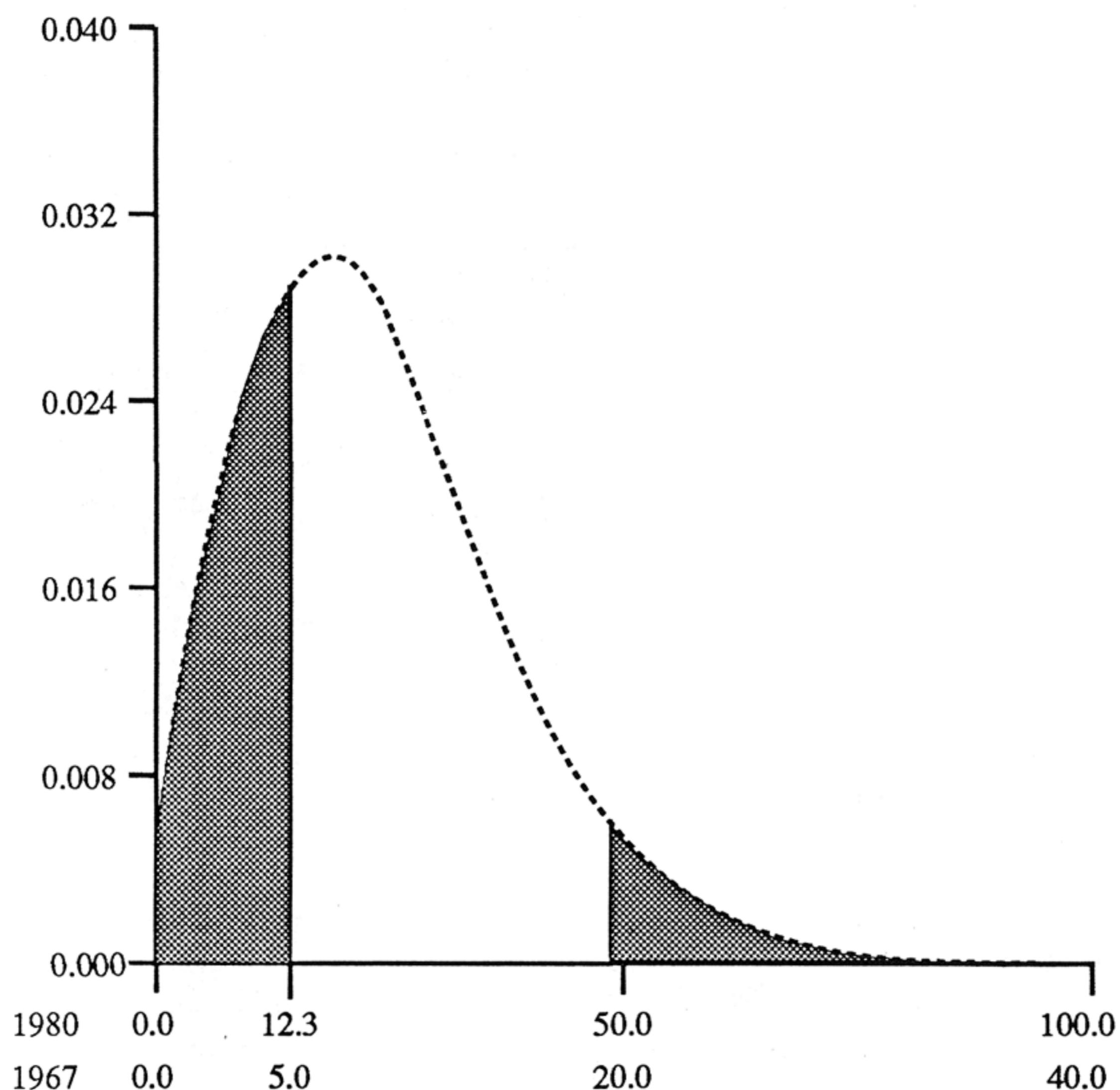
Richard Butler and I used the more general GB2 to fit family income data for the thirty-three-year period from 1948 to 1980.²³ The results in figure 18 demonstrate that the GB2 is much more flexible and provides a significantly better fit than the lognormal.

FIGURE 18
1980 Income, Lognormal, and GB2



Given these fitted distributions, measures of inequality and other characteristics can be easily investigated. In figure 19, for example, the shaded area on the left represents the fraction of families with incomes less than \$12,300 in 1980. The shaded area on the right denotes the fraction of families with incomes greater than \$50,000 in 1980.

FIGURE 19
1980 Income Distribution using GB2



Since we are looking at incomes for more than thirty years, it is important to adjust for inflation in order to represent real purchasing power. All incomes will be converted to 1967 dollars. These adjustments for 1980 are represented at the bottom of figure 20. Thus \$5000 in 1967 has approximately the same purchasing power as \$12,300 in 1980, and \$20,000 in 1967 is equivalent to approximately \$50,000 in 1980.

We now consider how these distributions have moved over the thirty-three-year period from 1948 to 1980. We consider all families, white families, and black families.

FIGURE 20
All Families' Income

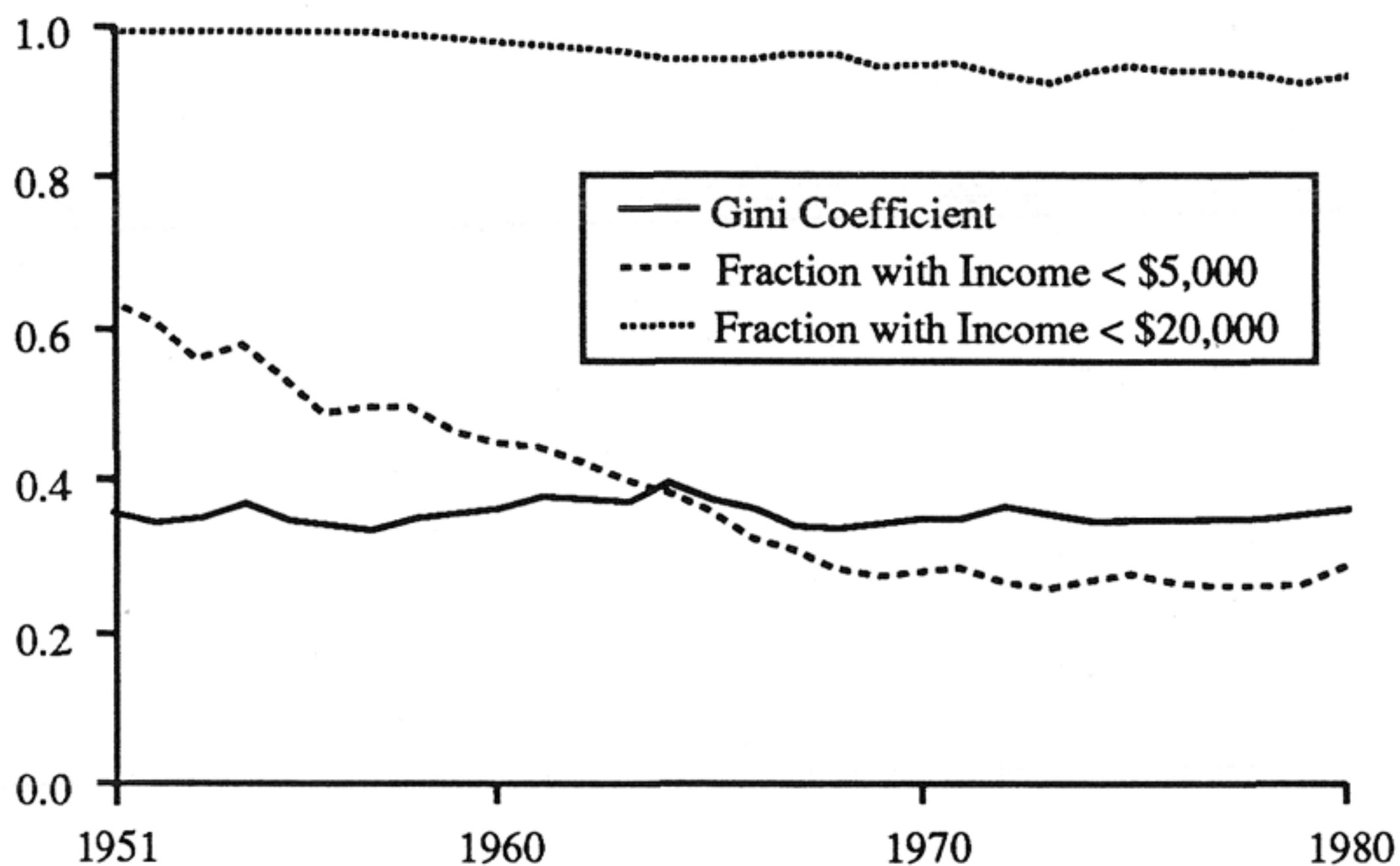


Figure 20 reports income characteristics for all families over this time period. The Gini coefficient is a measure of overall inequality. The small changes in this measure of inequality mask much larger changes in some other measures of economic well-being. For example, the fraction of the population with incomes less than \$5,000 (1967 dollars) has steadily decreased from 60 percent in 1948 to less than 30 percent in the late 1970s. Remember that these figures have been adjusted for inflation and represent real purchasing power. The fraction of the population with incomes less than \$20,000 (1967 dollars) has declined slightly from 99 percent to 95 percent in 1980. The area between the \$5,000 and \$20,000 lines represents a broad measure of the middle class. We see that the middle class with incomes between \$5,000 and \$20,000 has increased from 38 percent to 66 percent for this period.

It is interesting to compare the distribution of family incomes for blacks and whites.

FIGURE 21
Black Families' Income

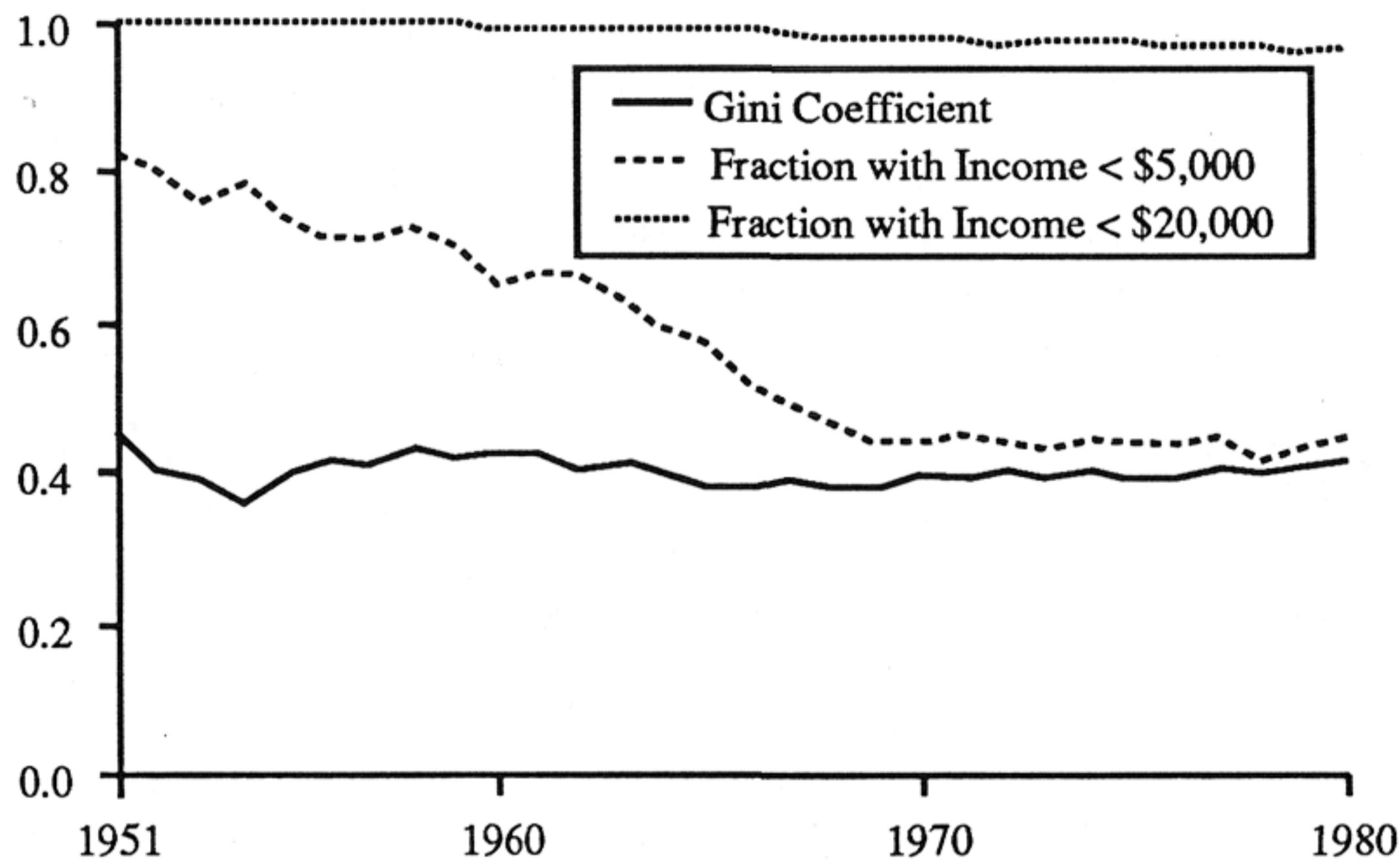
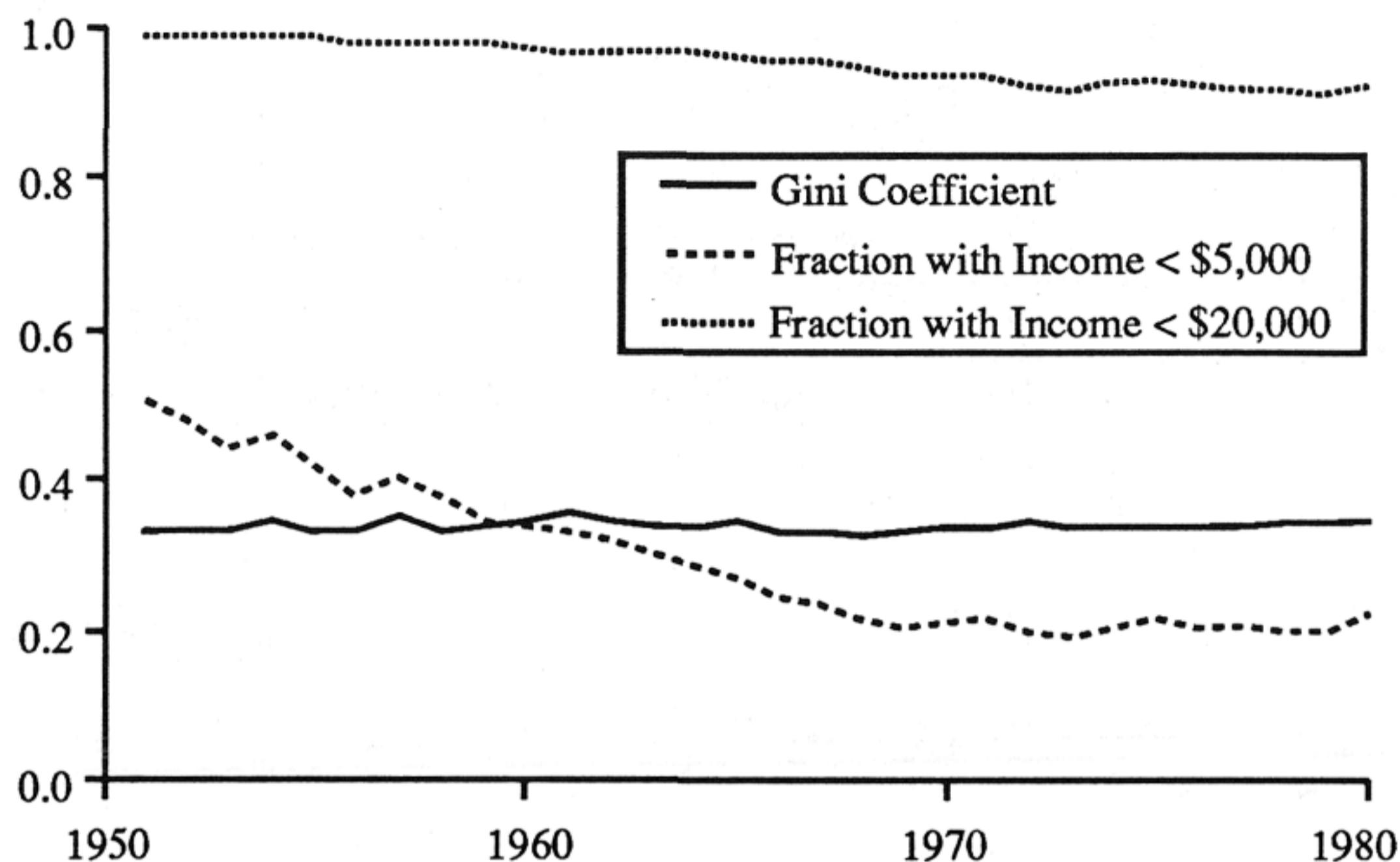


Figure 21 reports similar information for black families. We see that overall income inequality decreases slightly during the mid to late 1960s and then increases gradually. The Equal Employment Opportunity and Affirmative Action programs began in the mid 1960s. The fraction of black families with incomes less than \$5,000 (1967 dollars) has decreased dramatically over this period from 85 percent to 45 percent in 1980. Thus 55 percent of black families had incomes above the \$5,000 level by 1980.

FIGURE 22
White Families' Income



Income inequality for white families has been relatively constant over this time period. Again, there are changes in the distribution that this statistic does not reflect. White families with incomes less than \$5,000 (1967 dollars) decreased from 55 percent to 22 percent, and those with incomes greater than 20,000 (1967 dollars) increased from about 1 percent to 8 percent. Thus by 1980, 78 percent of white families had incomes greater than 5,000 (1967 dollars), compared with 55 percent of black families.

It should be apparent that there have been considerable movements of the distributions of income for blacks and whites over time, and both groups are better off. But how have these two distributions moved relative to each other?

FIGURE 23
Distance between Black and White Distributions

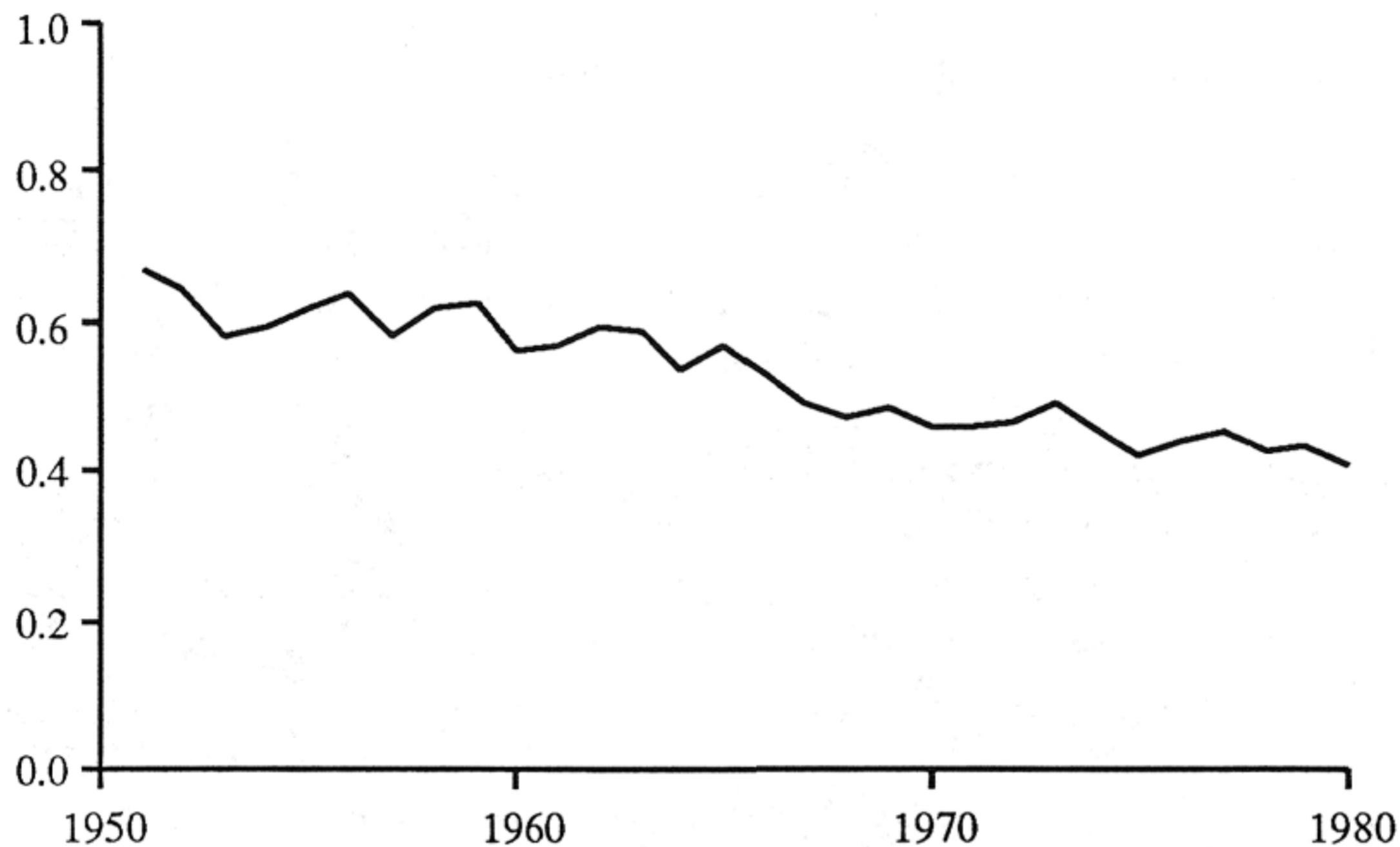


Figure 23 depicts a measure of the distance between the income distribution for white and black families. We observe that there are large reductions in the distance between the two distributions. This represents very large changes in the economic well-being of black families relative to white families over the entire time period. These changes predate the social legislation of the 1960s and continue through the 1970s. What factors seem to be associated with these movements?

- (1) Economic growth is an important factor and is associated with increased equality for blacks and whites. This appears to result from increasing the fraction of families with incomes greater than \$5,000 (1967 dollars) more than it increases the fraction of families with incomes greater than \$20,000 (1967 dollars). In other words, growth appears to be associated with everyone being better off, but relatively speaking the lower income families are helped the most.
- (2) On the other hand, Inflation tends to increase income inequality. Inflation decreases real income or purchasing power—especially for those with relatively fixed incomes. Inflation was seen to increase the fraction of families with incomes less than \$5,000 (1967 dollars).
- (3) Government expenditure, transfer payments, and equal employment opportunity legislation appeared to have little impact on the distribution of income for whites, but it did help shift blacks above the \$5,000 level.

In summary, we have found important changes in the distribution of income over time with a narrowing of the disparity between the

income distributions for black and white families. Inflation and economic growth have an impact on the distribution of income. Government programs have been helpful in improving the economic well-being of blacks.

Distribution of Coal Particle Size

Distributions of the size of coal particles are important in coal combustion. I worked with Dale Richards, Philip Smith, and Bill Sowa in analyzing the distribution of sizes of pulverized coal.²⁴ This is related to a multi-million dollar research grant received by the Advanced Combustion Research Center at BYU. This project is directed by L. Douglas Smoot, who is investigating ways to make coal burn more efficiently. Pulverized coal has been used as a fuel for commercial combustion since the late 1800s and currently accounts for a major portion of the power generated by electric utilities. Pulverized coal combustion requires grinding coal into very small sizes and then mixing it with steam or oxygen in a combustion chamber. The mixture is burned, creating steam that generates electricity. The distribution of particle size is important to the efficiency and operation of the furnace. Small particle sizes are important to insure rapid ignition, and some larger particle sizes are needed to obtain maximum combustion efficiency. The Combustion Research Center has built computer simulation models (thirty thousand lines of Fortran code, fifteen CPU hours/case) to determine the relationship between the distribution of the size of coal particles, other inputs, and the electricity generated and related pollution.

The distribution of coal particle size is used in these computer simulation models. An accurate model of the distribution of particle size is needed. The lognormal has been one of the most widely used models to date. Since coal particle size distributions can have many possible shapes, the additional flexibility of the GB2 may be very useful over a wide range of conditions. In figures 24 and 25 we see two examples. In figure 24 the LN and the GB2 both provide a good fit. In figure 25 the GB2 provides a much better fit than the lognormal. The GB2 will always do at least as well as any of its special cases.

FIGURE 24
Combustion Coal (Wyoming) - HIST - GB2 - LN

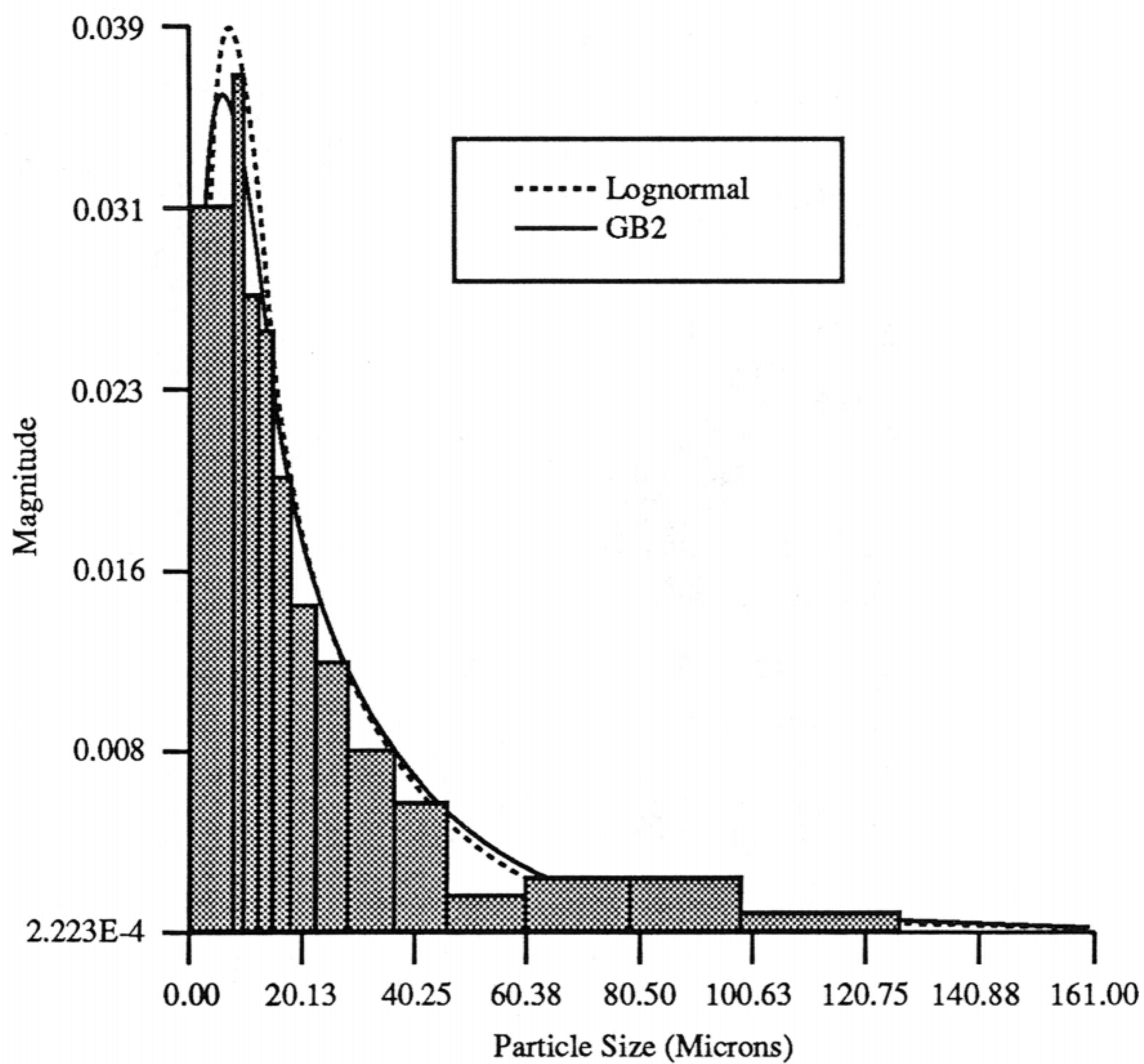
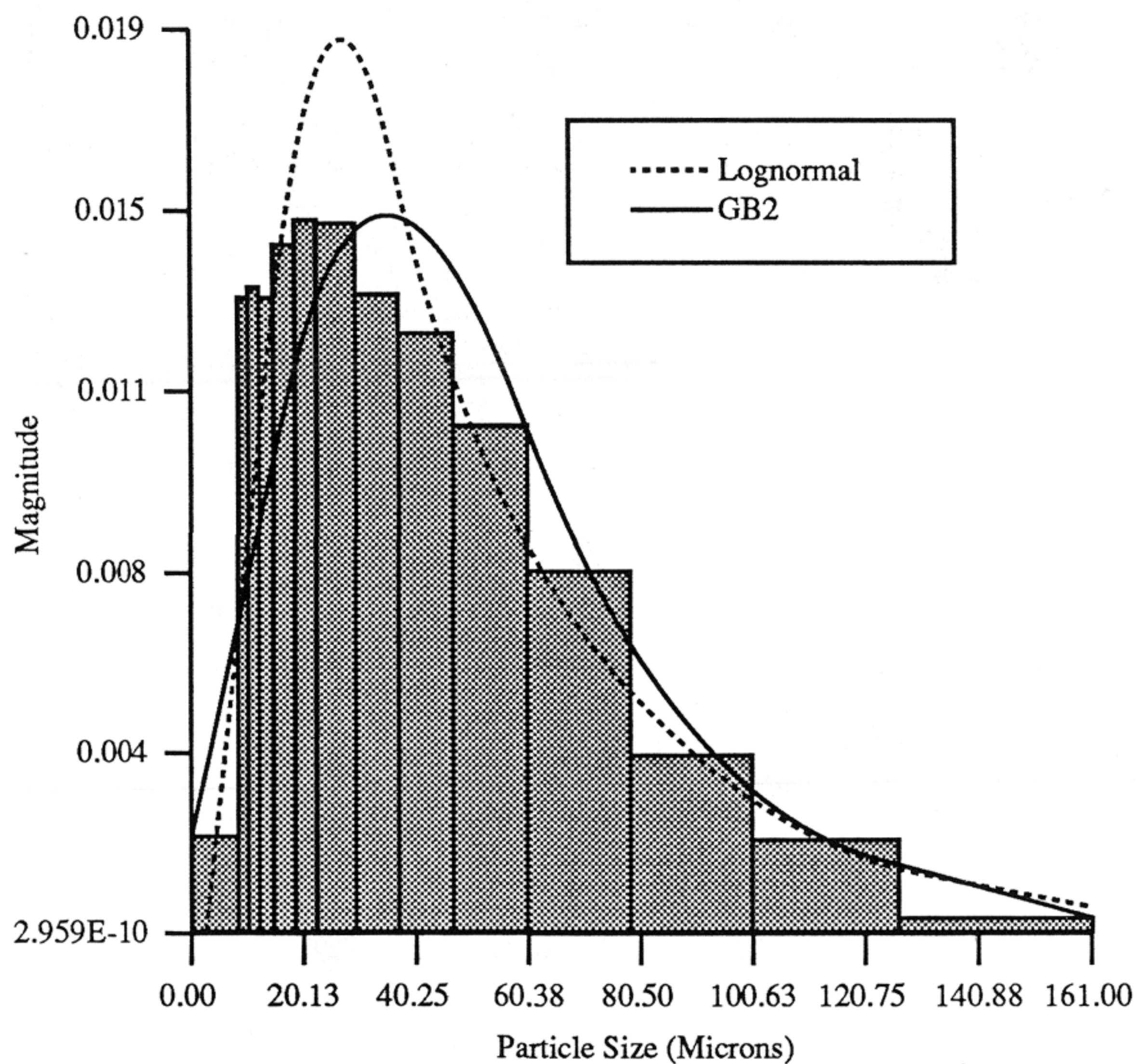


FIGURE 25
Gasification Coal (Utah) - HIST - GB2 - LN



The results of this study suggest that the distribution does matter. Furthermore, it may be possible using this methodology to help determine optimum particle distributions.

Distribution of Stock Prices and Returns

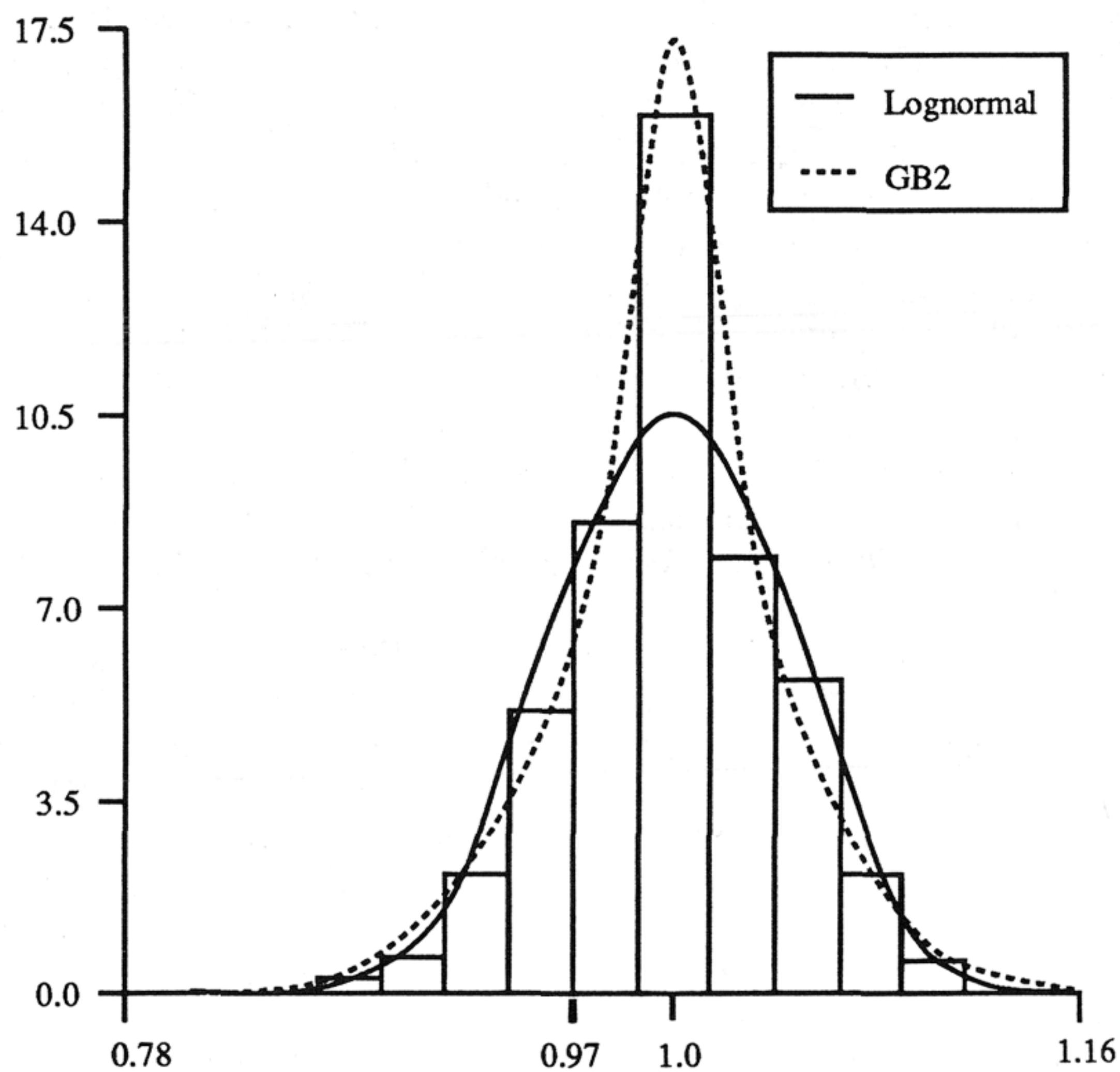
The last application deals the distribution of stock prices. The form of the distribution of returns on securities and portfolios is important for several reasons. The distribution of returns or profits or losses on a security determines the expected or average return as well as reflecting the risk in the investment. The probabilities of large deviations from the mean may be much different for one security than for another. These factors are of major interest and concern to brokerage firms and those with investment responsibilities. The world of finance has simultaneously become more complicated and exciting with the introduction of new financial instruments. Options are commonplace, as are terms such as puts and calls, hedges and stop-loss orders, and options on stock market indices. An investor in this new environment is still faced with the assessment of unknown probabilities about the likelihood of the future price of a security or other financial instrument increasing or decreasing by a certain amount. The famous Black-Scholes option pricing formula is an example of an effort to assess these probabilities. In order to do so, it is important to be able to accurately describe the shape of the distribution of prices or returns.

Price changes are often assumed to be distributed as a normal or lognormal.²⁵ A number of studies have shown that daily stock returns have distributions that are more peaked than the normal or lognormal and also assign higher probabilities of large returns or losses than the normal.²⁶ In other words, the tails of the normal or lognormal are not thick enough. Studies of monthly returns suggest distributions that are slightly skewed to the right. I am currently investigating these distributions in more detail and considering some related issues with Richard Bookstaber and Ray Nelson.

In studying the distribution of daily returns, the GB2 provides a much better fit than the lognormal in almost all of the cases considered. Figure 26 shows the distribution of daily returns for five hundred observations on the stock Compugraphic. The returns are calculated by dividing today's price by yesterday's price and will equal one if there is no change in the price. The returns are roughly centered around one, which means that on average the price changes are approximately zero. However, there are some large increases and decreases over the time period, as is reflected in the variation of the distribution, sometimes exceeding 15 percent in one day. The empirical data are seen to have a distribution that is more peaked near the mean and has thicker tails than

the lognormal. Recall that the lognormal is too peaked for the income data. The GB2 fits the data remarkably well throughout the entire range.²⁷

FIGURE 26
Stock Returns with Lognormal and GB2



We are also analyzing the distribution of seven years' of monthly data on approximately one thousand stocks listed on the New York Stock Exchange (CRSP tapes). The mean, variance, skewness, and a measure of thickness of the tails have been calculated for each of the stocks. Figure 27 contains a summary of these calculations.

FIGURE 27
Stock Returns: 1,000 Stocks

		SKEWNESS		
		Negative	Normal	Positive
		-	0	+
TAIL THICKNESS	Thin	0 %	0 %	0 %
	Normal	0 %	60 %	10 %
	Thick	1 %	9 %	20 %

The columns allow for positive or negative skewness as well as symmetric distributions. The rows correspond to different thickness of the tails with the "normal" providing the bench mark. The central block containing 60 percent corresponds to returns that are "roughly" normally distributed (+/- two standard deviations). There is a significant occurrence of distributions of returns that are thick tailed and skewed. Approximately 30 percent of the stocks have distributions that are significantly positively skewed, and about the same percentage have tails much thicker than the normal. The GB2 distribution provides a significantly better fit to the distributions of these returns than the normal or lognormal distributions.²⁸

These results provide strong evidence that there are models which provide a better fit to stock returns than the commonly used models. A number of methods of analysis used in finance are implicitly based upon assumptions of normality or lognormality of returns. These include

methods such as the Black-Scholes formula for determining the value of options as well as methods of estimating the risk of a stock or portfolio as represented by the betas. The results from both of these methods are sensitive to the underlying distribution of returns. Statistical distributions provide the basis for some exciting research in many areas in finance.

I have found statistical distributions and their various applications to be an exciting area to study. I believe this work has important applications in many areas. As I reflected on this, I came across a statement made by Francis Galton in the introduction to his book *Natural Inheritance*, which expresses my feelings about the topic and provides a fitting note to end on.

The road to be traveled over is full of interest of its own. It familiarizes us with the measurement of variability, and with the curious laws of chance that apply to a vast diversity of social subjects. This part of the inquiry may be said to run along a road on a high level, that affords wide views in unexpected directions and from which easy descents may be made to totally different goals to those we have now to reach. I have a great subject to write upon.²⁹

NOTES

¹M. G. Kendall, "Where Shall the History of Statistics Begin?" *Biometrika* 47 (1960): 447-49.

²F. N. David, "Dicing and Gaming [Note on the History of Probability]," *Biometrika* 42 (June 1955): 1-15; see also F. N. David, *Games, Gods and Gambling* (London: Charles Griffin, 1962).

³David, "Dicing and Gaming," 6.

⁴See, generally, Kendall, "Where Shall the History of Statistics Begin?"; and David, "Dicing and Gaming."

⁵Robin L. Plackett, "The Principle of the Arithmetic Mean," *Biometrika* 45 (1958): 130-35.

⁶David, "Dicing and Gaming," 7.

⁷Abraham M. Hasofer, "Random Mechanisms in Talmudic Literature," *Biometrika* 54 (June 1967): 316-21.

⁸M. G. Kendall, "The Beginnings of a Probability Calculus," *Biometrika* 43 (June 1956): 1-14.

⁹See, generally, Hasofer, "Random Mechanisms."

¹⁰See, generally, Ian Hacking, *The Emergence of Probability* (Cambridge: Cambridge University Press, 1975); and L. E. Maistrov, *Probability Theory: A Historical Sketch*, ed. and trans. Samuel Kotz (New York: Academic Press, 1974).

¹¹Edmund Halley, "An Estimate of the Degrees of Mortality of Mankind, Drawn from Curious Tables of the Births and Funerals at the City of Breslaw; with an Attempt to Ascertain the Price of Annuities upon Lives," *Philosophical Transactions of the Royal Society of London* 17 (1693): 596-610.

¹²See, generally, Abraham DeMoivre, *Annuities upon Lives* (London: W. Pearson, 1725).

¹³See, generally, Abraham DeMoivre, *Annuities upon Lives*, 2d ed. (London: H. and G. Woodval, 1743).

¹⁴E. S. Pearson, "Some Incidents in the Early History of Biometry and Statistics, 1890-94," *Biometrika* 52 (1965): 6.

¹⁵Ibid., 7.

¹⁶E. S. Pearson, "Some Reflexions on Continuity in the Development of Mathematical Statistics, 1885-1920," *Biometrika* 54 (1967): 343.

¹⁷Stephen M. Stigler, "Napoleonic Statistics: The Work of Laplace," *Biometrika* 62 (1975): 503.

¹⁸Quoted in J. Aitchison and J. A. C. Brown, *The Lognormal Distribution with Special References to Its Uses in Economics* (Cambridge: Cambridge University Press, 1969).

¹⁹Pearson, "Some Incidents," 10.

²⁰Kenneth J. White and Nancy G. Horsman, *Shazam: User's Reference Manual* (Vancouver: University of British Columbia, 1986), 5, 146, 163.

²¹For the GB2 see A. M. Mathai and R. K. Saxena, "On a Generalized Hypergeometric Distribution," *Metrika* 2 (1966): 127–32; R. L. Prentice, "Discrimination among Some Parametric Models," *Biometrika* 62 (1975): 607–14; and J. B. McDonald, "Some Generalized Functions for the Size Distribution of Income," *Econometrica* 52 (1984): 647–63. For the GB1 see McDonald, "Generalized Functions." For the generalized t (GT) see J. B. McDonald and W. Newey, "Partially Adaptive Estimation of Regression Models vis the Generalized t Distribution," forthcoming in *Econometric Theory*; and J. B. McDonald, "Partially Adaptive Estimation of ARMA Models," forthcoming in *International Journal of Forecasting*.

²²See, generally, Roger Joseph Boscovich and Christopher Maire, *De Litteraria Expeditone per Pontificiam ditionem ad dimetiendas duas Meridiani gradus* (Rome: Palladis, 1755); Roger Joseph Boscovich and Christopher Maire, *Voyage astronomique et geographique, dans l'etat de l'eglise* (Paris: N. M. Tillard, 1770); Churchill Eisenhart, "Boscovich and the Combination of Observations," chap. 7 of *R. J. Boscovich, Studies of His Life and Work*, ed. L. L. Shyte (London: Alen, 1961); Leonhard Euler, *Recherches sur la questions des inegalites du mouvement de Saturne et de Jupiter, sujet propose pour le prix de l'annee 1748, par l'Academie royale des sciences de Paris* (Basel: Turici, 1749); Carl Friedrich Gauss, *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae* (Gottingen: Dieterich, 1823); Carl Friedrich Gauss, *Theoria motus corporum celestium* (Hamburg: Perthes et Besser, 1809) [translated into English as *Theory of Motion of the Heavenly Bodies Moving about the Sun in Conic Sections*, trans. Charles Henry Davis (1857; reprint, New York: Dover Publications, 1963)]; Pierre Simon Laplace, "Theorie de Jupiter et de Saturne," *Memoires de l'Academie royale des sciences de Paris* (1785): 33–160; Robin L. Plackett, "A Historical Note on the Method of Least Squares," *Biometrika* 36 (December 1949): 458–60; Robin L. Plackett, "The Discovery of the Method of Least Squares," *Biometrika* 59 (July 1972): 239–51; and Hilary L. Seal, "The Historical Development of the Gauss Linear Model," *Biometrika* 54 (June 1967): 1–24.

²³Richard J. Butler and James B. McDonald, "Income Inequality in the U.S.: 1948–80," *Research in Labor Economics* 8 (1986): 85–140; see also James B. McDonald, "Some Generalized Functions for the Size Distribution of Income," *Econometrica* 52 (May 1984): 647–63.

²⁴J. B. McDonald et al., "Statistical Distributions of Coal Particle Sizes in Pulverized-Coal Combustion and Gasification," copy of manuscript in author's possession.

²⁵See, generally, Louis Bachelier, *Theory de la speculation* (Paris: Gauthier-Villars, 1900); M. F. M. Osborn, "Brownian Motion in the Stock Market," *Operations Research* 7 (January–February 1959): 100–128.

²⁶Benoit Mandelbrot, "The Variation of Certain Speculative Prices," *Journal of Business* 36 (October 1963): 394–419.

²⁷Richard Bookstaber and James McDonald, "A General Distribution for Describing Security Price Returns," *Journal of Business* 60 (July 1987): 401–24; James McDonald and D. O. Richards, "Model Selections: Some Generalized Distributions," *Communications in Statistics: Theory and Methods* 16, no. 4 (1987): 1049–74.

²⁸James McDonald and Ray Nelson, "Adaptive Beta Estimation for the Market Model" copy of manuscript in author's possession.

²⁹Thomas Galton, *Natural Inheritance* (London: Macmillan, 1890), 3.