THE MAXIMIZATION OF THE LOGARITHMIC ENTROPY FUNCTION

THE MAXIMIZATION OF THE LOGARITHMIC ENTROPY FUNCTION AS A NEW EFFECTIVE TOOL IN STATISTICAL MODELING AND ANALYTICAL DECISION MAKING

by

Yosri Diab, B.Sc., Mech. Eng.

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Engineering

McMaster University

April 1972

Master of Engineering, (1972)
(Mechanical Engineering)

McMaster University
Hamilton, Ontario

TITLE: THE MAXIMIZATION OF THE LOGARITHMIC ENTROPY FUNCTION AS A NEW EFFECTIVE TOOL IN STATISTICAL MODELING AND ANALYTICAL DECISION MAKING

AUTHOR: YOSRI DIAB, B.Sc., Mech. Eng. (Alexandria)

SUPERVISOR: Professor J. N. Siddall

ABSTRACT

This thesis introduces a new effective method in statistical modeling and probabilistic decision making problems. The method is based on maximizing the Shannon Logarithmic Entropy Function for information, subject to the given prior information to serve as constraints, to generate a probability distribution. The method is known as the Maximum Entropy Principle or "Jaynes Principle". Tribus used it earlier, but in a limited case, without general application to either statistical modeling or probablistic decision making. In this thesis, a new method which generalizes the above principle is introduced. This permits practical applications, some of which are illustrated.

ACKNOWLEDGMENTS

For contributing to the completion of this thesis I should like to express my personal thanks and appreciation to

Professor J. N. Siddall, the supervising professor, for his continued interest, patience and suggestions.

Richie Pepin and Dave Stratton for their help and encouragement.

The McMaster Computer Centre for providing the computing facilities.

This work was supported by a grant from the National Research Council of Canada.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

In a decision process, once the problem is specified mathematically, part of the basis for a choice between different strategies or designs lies in the prior probabilities or probabilities of relevant events. To satisfy this logical foundation of decision theory, the problem of formulating prior probabilities should be based on a rigorous principle.

Shannon's Entropy Function[10] was one of the first steps in this direction. Later Jaynes[12] introduced a principle (known as the Maximum Entropy Principle), which is based on maximizing Shannon's Entropy Function, subject to the given information as constraints. Although the principle is one of the most important steps towards formulating prior probabilities, it was difficult to obtain a general solution for a general problem. Tribus [2] expressed this principle mathematically, but his applications were limited to certain states of knowledge, of prior information. He did not provide a general solution. Some of the resulting distributions were the gamma, the exponential, and the normal distributions. Obviously these curves cannot provide an adequate representation of many of the distributions encountered in statistical practice.

In this thesis a new algorithm is introduced using the Maximum Entropy Principle to generate a general probability distribution from the first moments. The old problem of representing data by using the first four moments has been solved by the new algorithm. A comparison has been made between the empirical existing methods and the new algorithm in order to show the power of the principle. Also, in analytical decision theory, the problem of predicting the

probability density functions of a random variable when this random variable is a function of many other known random variables has been solved using the new algorithm.

In both applications (moment generation or in analog prediction under risk) a computer program has been written in FORTRAN IV language.

This thesis is mainly concerned with the generalization of the principle based on the first moments, but other types of functions could be used rather than the moment function (a slight modification on the algorithm would be needed).

In Chapter II, a very brief introduction to analytical decision theory is given to shed some light on the area where the new algorithm could be applied, and to show what methods in this area exist.

CHAPTER II

A BRIEF INTRODUCTION TO ANALYTICAL DECISION THEORY

2.1 DECISION AND DECISION THEORY

A decision is a selection, which involves risk, between alternative actions. For a decision to be possible, there must be two or more alternatives available. These alternatives represent a set of possible acts which the decision maker may choose. The acts are connected in some way to result in a set of possible outcomes. If there is a known deterministic connection between the acts and outcomes, the problem is one of deterministic choice. If the decision maker knew which outcome would result from each act, he could choose the act which resulted in the outcome he most valued. The choice among outcomes reflects a value judgment. In other words the decision maker must know the values he associates with the various outcomes which may result from his choice of acts. There is thus no problem if there are no uncertainties.

Decision theory is, then, concerned with the making of decisions, i.e. choice of acts, in the face of uncertainty. The uncertainty may be concerned with the relation between acts and outcomes or it may be related to reliability of the available information. The maximization of the logarithmic entropy function represents part of an attempt to provide a rational basis for decision making under uncertainty.

2.2 THE DECISION PROCESS

Consider Figure 2.1, in which the several elements which enter a decision process are shown. Box 2 (Probability assignments or statistical modeling) serves to put together the prior general information and special evidence which

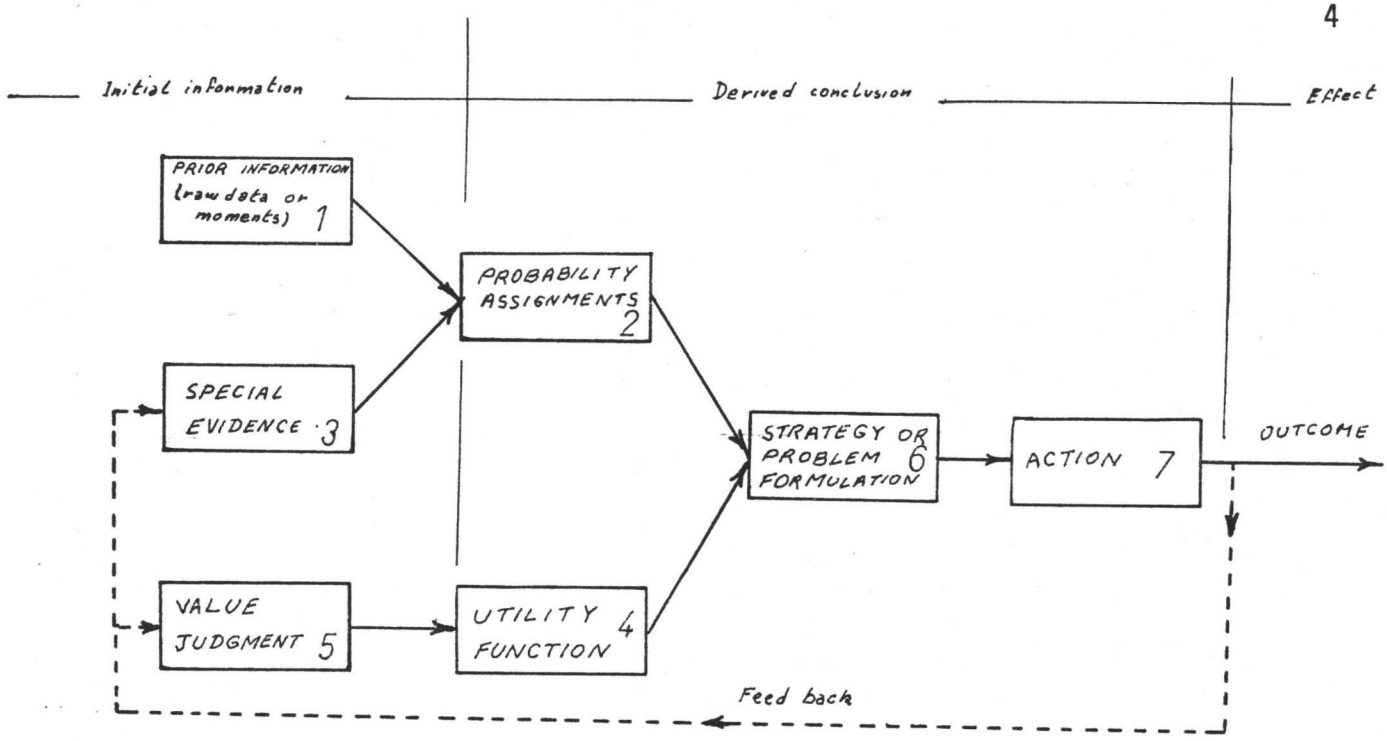


Fig. 2.1 The elements of a decision process.

pertains to the special case being treated, box 6 (strategy or problem formulation) serves to put together the probability assignments and the utility functions, the output of this box is a strategy for action, i.e. the basis for a decision. Once the problem is formulated, the action may be taken. The result of the action is to provide more data which may be used later on, assuming a second chance occurs. As it is shown the action or the decision depends entirely on the strategy formulation; a "bad" strategy would lead to a "bad" decision, and a "good" strategy would lead us to a "good" decision. In other words the strategy formulation is a method (or strategy) to combine the given probability assignment with the utility function. So there are three elements involved in the process, the utility function, the probability assignment, and the strategy. A "good" decision depends on all of them.

2.3 THE UTILITY FUNCTION

The utility function is sometimes called the value function or the loss function, and sometimes given a negative sign (the sign depends on whether you are normally pessimistic or optimistic). The utility function will not be discussed in detail as it is outside the scope of this thesis; the reader may refer to references (1) and (2).

Briefly, the utility function expresses or defines the value in a problem mathematically. A decision maker should know how to define the utility function and know how to discriminate among the outcomes. Decision theory will be of no help to a decision maker who does not know how to determine what he wants from what he does not want.

2.4 STATISTICAL MODELING (PROBABILITY ASSIGNMENT)

To proceed in a Decision Process (see Figure 2.1), some or all elements of the utility function should be defined in probability terms to the best of prior knowledge.

A prior knowledge about a random variable varies from one fully defined to one that is undefined, but in all cases it is known that a distribution exists.

2.4.1 Degrees of Prior Knowledge

The degrees of prior knowledge could be classified as follows:

a) The exact prior distribution is known.

b) The first m moments of the prior distribution are known (m = 1, 2, ...), i.e. the actual numbers are available for calculation.

c) Sample data is available.

d) No prior knowledge other than the existence of a prior distribution is available.

Usually, prior knowledge is limited to cases (b) and (c), where some independent observations (or raw data), or some of the moments, are available.

With knowledge as given in case (c), and sometimes as in case (b), it is difficult to proceed in the decision process. The available information should be in a mathematical form or other suitable form. A suitable mathematical model must therefore be found to describe the prior information. This suitable model could be one of the well known analytical distributions such as the normal or Weibull, or generated by one of the empirical approximation methods like the Johnson, the Pearson, the Cornish-Fisher expansion, the Gram-Charlier series, the Edgeworth series[3], or the most recent method, the maximum-logarithmic entropy distribution method, which is introduced in this thesis. Since this new method will be applied to statistical modeling, a brief review of some of the above methods will be given in the following pages for comparison.

2.4.2 Analytical Distributions

The normal or Gaussian distribution is the best known statistical model. However, many phenomena cannot be adequately described by a normal distribution. Other models like the gamma, the beta, the Chi-Square, the exponential, the uniform, the log-normal, the Rayleigh, the Cauchy, the Weibull[4] distributions, etc., could be successful models in describing specific phenomena but, generally, these analytical distributions do not describe accurately most phenomena.

Although some of these models do lead to a wide diversity of distribution shapes, they still do not provide the degree of generality that is frequently desirable. This is illustrated by Figure 2.2. This chart shows the regions in the (β_1 and β_2) plane where various analytical distributions can be fitted, where β_1 and β_2 are the square of the standardized measure of skewness

and the standardized measure of peakedness respectively. Distributions shown include the normal, beta (uniform special case), gamma (exponential special case), the log normal, and the student t distribution (a symmetric distribution that approaches the normal as its degree of freedom becomes arbitrarily large). All normal distributions ($\beta_1 = 0$ and $\beta_2 = 3$) are represented in Figure 2.2 by a single point; as are also the exponential and uniform distributions. The gamma and the log-normal distributions can be fitted for all value of β_1 and β_2 that fall on the curve shown near the centre of the chart. The beta distribution occupies a region in Figure 2.2, and thus provides greater generality than any of the other distributions.

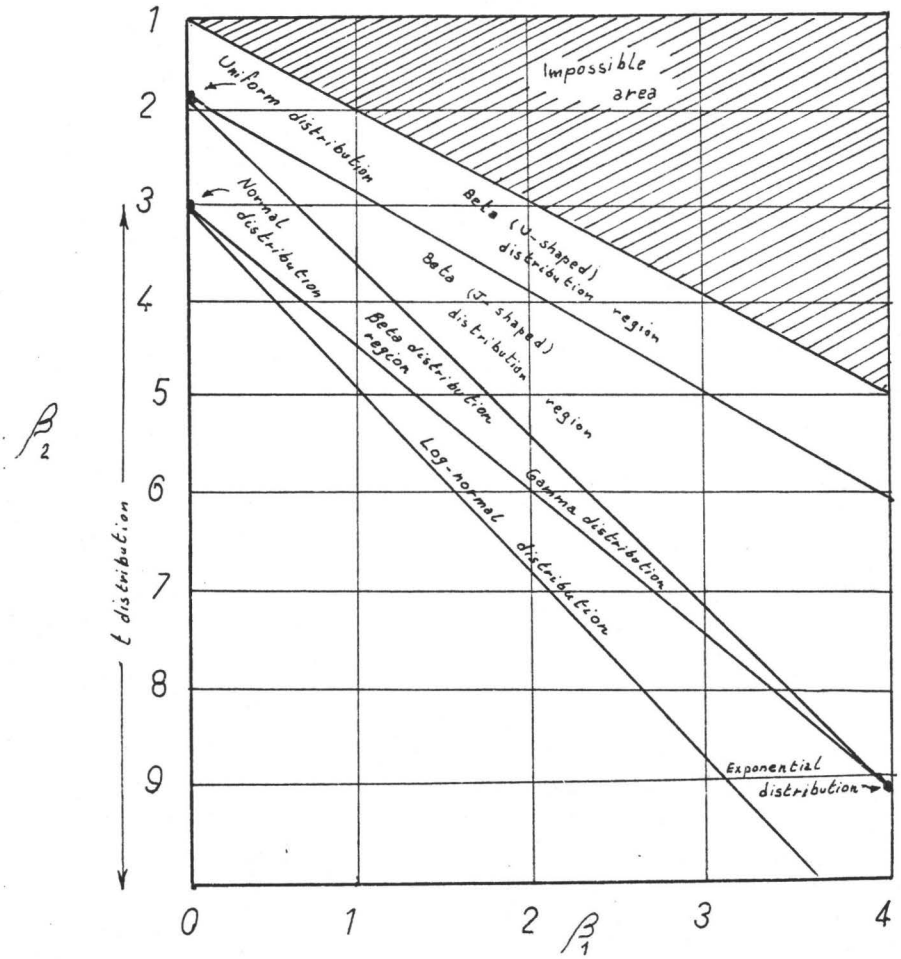


Fig. 2.2 Region in (β_1, β_2) plane for various distributions[4]

Note that there is a large region of values of β_1 and β_2 that is not covered by any of the above distributions.

2.4.3 Empirical Distributions

The main advantage of the empirical distributions over the analytical ones is the flexibility in covering a larger region. However for some regions, the empirical distributions deviate from the actual ones by a significant amount. Other regions are not covered by any distributions. Two main empirical distributions, Johnson's and Pearson's, will be discussed very briefly. For more details and for details on other methods (like Cornish-Fisher expansion, Gram-Charlier series and Edgeworth series) the reader can consult reference (3).

Johnson distribution

Johnson proposed empirical distributions based on the transformation of a standard normal variate. An advantage of such a transformation is that estimates of the percentiles of the fitted distribution can be obtained using a table of areas under a standard normal distribution. A disadvantage is that at least three points must be known to determine the final distributions, a method which limits itself to statistical modeling only. For more details the reader can refer to reference (5).

Pearson distribution

Karl Pearson proposed a group of distribution families. Each family can be generated as a solution to the differential equation

$$\frac{df(x)}{dx} = \frac{(x - \phi_3)\, f(x)}{\phi_0 + \phi_1 x + \phi_2 x^2}$$

where x is the random variable with probability density function f(x), and the ϕ's are parameters defining a specific distribution. The solution of this equation leads to a large number of distribution families. The descriptions of the procedure for fitting Pearson distributions to data are lengthy, since each family requires solution of a different set of equations. The underlying principles are reviewed in reference (3) and the formulae for each family are given in reference (7). Reference (6) includes tables for Pearson's functions, and a discussion of procedures for using the tabulations to obtain percentiles other than those tabulated. It indicates their possible use for the inverse problem of estimating accumulative probabilities corresponding to specified values of the random variables.

2.4.4 The Maximum-Logarithmic Entropy Distribution

The following section and chapters deal with this method. The theory, the algorithm, the application, and a comparison with other methods is discussed in detail.

2.5 FORMULATION OF STRATEGY OR PROBLEM

In general, the problem could be summarized as the determination of the distribution of a random variable y which is a known function of n random variables $x_1, x_2, \ldots x_n$. We may express this relationship as

$$y = g(x_1, x_2, \ldots x_n)$$

where the random variables are defined by their density functions, or by some of their lower moments. To the author's knowledge, three methods are known; the transformation of variables technique, the Monte-Carlo Simulation, and the

generation of system moments.

2.5.1 The Transformation of Variable Technique

This method is applicable to finding the distribution of simple functions of independent random variables. The method is practicable for relatively simple situations. A fairly complicated relationship may possibly be built up by a series of steps using three simple relations (multiplication, division, and addition) between two random variables[1]. The method is a very powerful technique, but only for independent variables.

2.5.2 Monte-Carlo Simulation

The method is based on the Monte-Carlo approach, in which actual experiments to statistically define the required distribution are simulated numerically. It is applicable to dependent and independent random variables. Although the method is very accurate when the sample size is very large, it is expensive in computation time in comparison with other methods. More details of the method are available in reference (9).

2.5.3 Generation of System Moments

For the general relation

$$y = g(x_1, x_2, \ldots, x_n)$$

Siddall[1] shows that it is possible to approximate the moments of y in terms of the moments of x_i's by using a truncated Taylor's series expansion about the expected values of the x_i's. The approximate moments are

$$C_{1y} = g(u_1,u_2,\ldots,u_n) + \frac{1}{2}\left\{ \sum_{i=1}^{n} \frac{\partial^2 g(u_1,u_2,\ldots,u_n)}{\partial x_i^2}\,\sigma_i^2 \right. \tag{2.1}$$

$$\left. + \; 2 \sum_{\substack{i \\ i<j}} \sum_{j} \frac{\partial^2 g(u_1,u_2,\ldots,u_n)}{\partial x_i \partial x_j}\,\sigma_{ij}^2 \right\} + \ldots$$

$$C_{2y} = \sum_{i=1}^{n} \left[\frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i} \right]^2 \sigma_i^2$$

$$+ \; 2 \sum_{\substack{i \\ i<j}} \sum_{j} \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i} \; \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_j} \; \sigma_{ij}^2$$

$$+ \sum_{i=1}^{n} \frac{\partial g(u_1,u_2,\ldots u_n)}{\partial x_i} \; \frac{\partial^2 g(u_1,u_2,\ldots,u_n)}{\partial x_i^2} \; c_{3i}$$

$$+ \sum_{\substack{i \\ i \neq j}} \sum_{j} \frac{\partial g(u_1,u_2,\ldots u_n)}{\partial x_i} \; \frac{\partial^2 g(u_1,u_2,\ldots,u_n)}{\partial x_j^2}$$

$$x \;\; E[(x_i-u_i)(x_j-u_j)^2]$$

$$+ \; 2 \sum_{\substack{i \\ i \neq j}} \sum_{j} \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i} \; \frac{\partial^2 g(u_1,u_2,\ldots,u_n)}{\partial x_i \, \partial x_j}$$

$$x \quad E[(x_i-u_i)^2(x_j-u_j)]$$

$$+ \quad 2 \sum_{\substack{i \\ i \neq j \neq k}} \sum_{j} \sum_{k} \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i} \frac{\partial^2 g(u_1,u_2,\ldots,u_n)}{\partial x_j \partial x_k}$$

$$C \; x \quad E[(x_i-u_i)(x_j - u_j)(x_k - u_k)] \tag{2.2}$$

where c_{3i} is the third central moment of x_i, $E(z)$ is the expected value of z.

If the x_i's are independent, all terms but the first and third drop out.

$$c_{3y} = \sum_{i=1}^{n} \left[\frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i} \right]^3 c_{3i}$$

$$+ \quad 3 \sum_{\substack{i \\ i \neq j}} \sum_{j} \left[\frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i} \right]^2 \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_j}$$

$$x \quad E[(x_i-u_i)^2(x_j-u_j)]$$

$$+ \quad 6 \sum_{\substack{i \\ i<}} \sum_{\substack{j \\ j<}} \sum_{k} \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i} \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_j}$$

$$x \quad \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_k} \; E[(x_i-u_i)((x_j-u_j)(x_k-u_k)] \tag{2.3}$$

If the x_i's are independent, all terms but the first drop out.

$$c_{4y} = \sum_{i=1}^{n} \left[\frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i} \right]^4 c_{4i}$$

$$+ 4 \sum_{\substack{i \\ i \neq j}} \sum_{j} \left[\frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i} \right]^3 \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_j}$$

$$\times \; E[(x_i-u_i)^3(x_j-u_j)]$$

$$+ 6 \sum_{\substack{i \\ i < j}} \sum_{j} \left[\frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i} \right]^2 \left[\frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_j} \right]^2$$

$$\times \; E[(x_i-u_i)^2 (x_j-u_j)^2]$$

$$+ 12 \sum_{\substack{i \\ i \neq j \neq k}} \sum_{j} \sum_{k} \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i}^2 \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_j} \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_k}$$

$$\times \; E[(x_i-u_i)^2 (x_j-u_j) (x_k-u_k)]$$

$$+ 24 \sum_{\substack{i \\ i < j < k < \ell}} \sum_{j} \sum_{k} \sum_{\ell} \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_i} \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_j} \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_k}$$

$$\times \; \frac{\partial g(u_1,u_2,\ldots,u_n)}{\partial x_\ell} \; E[(x_i-u_i)(x_j-u_j)(x_k-u_k)(x_\ell-u_\ell)] \qquad (2.4)$$

If the x_i's are independent, all terms but the first and third drop out. The fifth and sixth moment could be obtained by a similar approach, but for simplicity, the first four moments are considered accurate enough to describe any practical distribution. If more accuracy is required, the fifth, the sixth or the seventh moments should be considered. Using the above expressions, the first four moments of y can be obtained, if we know the first four moments of the x's. The first four moments can be calculated from statistical data, if necessary.

The next step is to generate the probability distribution of y in terms of its first four moments. The Johnson method cannot generate this distribution as it requires at least three percentiles to match, which in this case are unknown. The Pearson method and the Maximum Logarithmic Entropy method can be applied to generate the distribution, but the accuracy of each method is different. This will be considered in Chapter VII and Chapter VIII.

CHAPTER III

MAXIMIZATION OF THE LOGARITHMIC ENTROPY FUNCTION

3.1 DECISION, CHOICE AND CERTAINTY[10]

Suppose there is a set of n possible events whose probabilities of occurrence are $p_1, p_2, \ldots p_n$. These probabilities are all that is known concerning which event will occur. Is there any measure of how much "choice" is involved in the selection of the event to make the decision, or of how certain we are of the outcome? If such a measure, S, is established it should satisfy three conditions. The three conditions (given by Shannon)[10] are:

1) S should be continuous in the p_i's.
2) If all p_i are equal, $p_i = \frac{1}{n}$, then S should be a monotonic increasing function of n. With equally likely events the amount of choice or uncertainty about the outcome increases with the number of possible events.
3) If a choice is to be broken down into two successive choices, the original S should be the weighted sum of the individual values of S. The meaning of this is illustrated in Figure 3.1. At the left we have four possibilities, for which $p_1 = \frac{1}{8}$, $p_2 = \frac{3}{8}$, $p_3 = \frac{1}{3}$, $p_4 = \frac{1}{6}$.

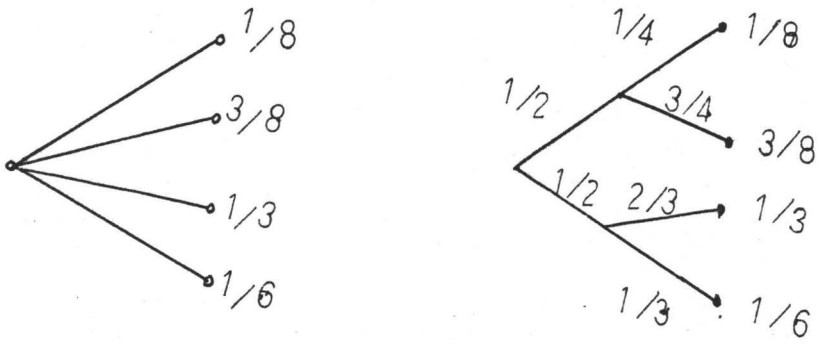


Figure 3.1 Decomposition of a choice from four possibilities.

On the right, we first choose between two possibilities, each with probability $\frac{1}{2}$; and if the first occurs we make another choice with probabilities $\frac{1}{4}, \frac{3}{4}$; and if the second occurs we make another choice with probabilities $\frac{2}{3}, \frac{1}{3}$. The final results have the same probabilities as before. We require, in this case, that:

$$S(\tfrac{1}{8}, \tfrac{3}{8}, \tfrac{1}{3}, \tfrac{1}{6}) = S(\tfrac{1}{2}, \tfrac{1}{2}) + \tfrac{1}{2}\, S(\tfrac{1}{4}, \tfrac{3}{4}) + \tfrac{1}{2}\, S(\tfrac{2}{3}, \tfrac{1}{3})$$

The coefficient $\frac{1}{2}$ is due to the second choice occurring only half the time.

3.2 THE LOGARITHMIC-ENTROPY FUNCTION

Shannon showed that the only S satisfying the three above assumptions(*) is of the form

$$S = -k \sum_i p_i \ln p_i \qquad (3.1)$$

The measure S is called the entropy. It has had a long and involved history. The word was originally coined from the Greek by Clausius in 1850 to mean transformation(**).

The entropy, S, has a number of interesting properties which further substantiate it as a measure of choice or certainty. Some of these properties are:

1) $S = 0$ if and only if all the p_i but one are zero, this one having the value unity. Thus, only when we are certain of the outcome, does S vanish. Otherwise S is positive.

(*) For the derivation see Appendix B.

(**) We suggest reference (11), for the reader who is interested in knowing the relation between this entropy and the one used in thermodynamics.

2) For a given n, S is a maximum and equal to k log n when all the p_i are equal (i.e. $p_i = \frac{1}{n}$). This is also intuitively the most uncertain situation.

3) Any change toward equalization of the probabilities $p_1, p_2, \ldots p_n$ increases S. Thus, if $p_1 < p_2$ and we increase p_1, decreasing p_2 an equal amount so that p_1 and p_2 are more nearly equal, then S increases. More generally, if we perform any "averaging" operation on the p_i of the form

$$\tilde{p}_i = \sum_j a_{ij} p_j$$

where $\sum_i a_{ij} = \sum_j a_{ji} = 1$, and all $a_{ij} \geq 0$,

then S increases (except in the special case where this transformation amounts to no more than a permutation of the p_j, when S of course remains the same).

The above properties are due to Shannon[10]. An additional important property was discovered by Jaynes[12]. Its direct application is known as the Maximum-Entropy Principle.

4) For a given n, when all or some of the event probabilities are subject to constraints, or relations between each other, and if all p_i's are given unbiased values, then S is MAXIMUM. This is the most uncertain situation for the specified constraints. The principle can be used to choose an unbiased set of p_i's consistent with known information about them. In case there are no constraints imposed on the problem, all p_i's should be equal, then S is maximum when it is equal to k log n, (which is the property (2)).

3.3 THE MAXIMIZATION OF THE ENTROPY FUNCTION

We are interested in the case when prior information is available, and an unbiased distribution is required. The use of property (4) is known as the "Maximum Entropy Principle", first introduced by E. T. Jaynes[2, 13]. The following statement was put forward by him. "The minimally prejudiced probability distribution is that which maximizes the entropy(*) subject to constraints supplied by the given information."

3.4 MAXIMIZING THE SO-CALLED GENERAL ENTROPY FUNCTION

Although the Shannon derivation of the logarithmic entropy function (Appendix B) demonstrates a convincing validity for the expression $p \ln p$, considered the so-called "General Entropy Function" introduced by Behara and Nath[14] which they say "in special cases, reduces to the Shannon entropy function". The function has the form

$$S_{\alpha} = \frac{p - p^{\alpha}}{1 - 2^{1-\alpha}} \tag{3.2}$$

where

$$\alpha \in (0, \infty)$$

Now let us apply condition (3) which is--"If a choice be broken down into two successive choices, the original S should be the weighted sum of the individual values of S." Assume a choice p_c is broken into p_d, p_e and the results are p_a, p_b. This is illustrated in Figure 3.2.

(*) Jaynes meant Shannon's entropy function ($p \ln p$).

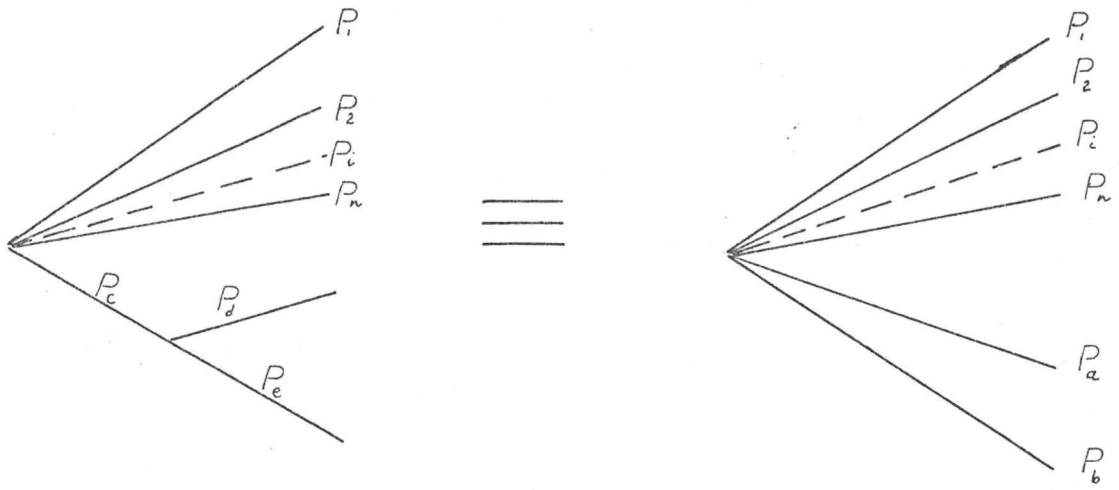


Fig.32 Successive choices.

It is clear that $p_a = p_c.p_d$ and $p_b = p_c.p_e$. By condition (3),

$$S(p_1, p_2, \ldots p_n, p_a, p_b) = S(p_1, p_2, \ldots p_n, p_c) + p_c\, S(p_d, p_e)$$

$$S(p_1, p_2, \ldots p_n) + S(p_a, p_b) = S(p_1, p_2, \ldots p_n) + S(p_c) + p_c\, S(p_d, p_e)$$

$$S(p_a, p_d) = S(p_c) + p_c S(p_d, p_e) \qquad (3.3)$$

Substituting (3.2) in (3.3) gives

$$\frac{p_a - p_a^{\alpha}}{1 - 2^{1-\alpha}} + \frac{p_b - p_b^{\alpha}}{1 - 2^{1-\alpha}} = \frac{p_c - p_c^{\alpha}}{1 - 2^{1-\alpha}} + p_c \left[\frac{p_d - p_d^{\alpha}}{1 - 2^{1-\alpha}} + \frac{p_e - p_e^{\alpha}}{1 - 2^{1-\alpha}} \right]$$

We multiply both sides by $[1 - 2^{(1-\alpha)}]$ and substitute $p_a = p_c\, p_d$ and $p_b = p_c\, p_e$

$$p_c\, p_d - p_c^{\alpha} p_d^{\alpha} + p_c\, p_e - p_c^{\alpha} p_e^{\alpha}$$

$$= p_c - p_c^{\alpha} + p_c\, p_d - p_c\, p_d^{\alpha} + p_c\, p_e - p_c\, p_e^{\alpha}$$

$$-p_c + p_c\, p_d^{\alpha} + p_c\, p_e^{\alpha} = -p_c^{\alpha} + p_c^{\alpha} p_d^{\alpha} + p_c^{\alpha} p_e^{\alpha}$$

$$p_c(p_d^{\alpha} + p_e^{\alpha} - 1) = p_c^{\alpha}(p_d^{\alpha} + p_e^{\alpha} - 1)$$

Diving both sides by $p_d^{\alpha} + p_e^{\alpha} - 1$

$$p_c = p_c^{\alpha} \tag{3.4}$$

To satisfy this relation α should be equal to 1. However, for α equal to 1, the so-called general entropy function reduces to the logarithmic entropy function, $-p \ln p$. So the general entropy function does not appear to be valid except in the special case where it reduces to the Shannon-Entropy Function case.

CHAPTER IV

THE MATHEMATICAL FORMULATION OF THE MAXIMUM LOGORITHMIC ENTROPY DISTRIBUTION

4.1 GENERAL FORMULATION OF THE EXPRESSION DEFINING THE BEST ESTIMATE OF PROBABILITIES

Quite often the information available from estimating probabilities is in the form of averages of certain functions, which we shall designate $f_r(x)$. That is, the form of the functions is known but all that is given is the mean value $< f_r(x) >$, for each of the functions $f_1(x)$, $f_2(x)$, . . . $f_r(x)$. A probability distribution must be generated which agrees with these averages but is maximally-non-committed with respect to anything else. The problem may be stated mathematically as follows. Maximize

$$S = -K \sum_{i=1}^{m} p_i \ln p_i \tag{4.1}$$

where

$$p_i = P\,(x_i \mid < f_1(x) > \; < f_2(x) > \; . \; . \; . \; < f_n(x) >) \tag{4.2}$$

x_i = the ith value of x

$< f_1(x) >$ = the mean value of $f_1(x)$

: =

: =

: = ...

:

$< f_n(x) >$ = the mean value of $f_n(x)$

m = number of events

Subject to the constraints

$$\sum_i p_i = 1 \tag{4.3}$$

$$\sum_i p_i f_1(x_i) = < f_1 >$$

$$\sum_i p_i f_2(x_i) = < f_2 >$$

:

:

:

$$\sum_i p_i f_n(x_i) = < f_n > \tag{4.4}$$

Expression (4.4) can be written in the compact form

$$\sum_i p_i f_j(x_i) = < f_j > \quad j = 1, 2, \ldots, n \tag{4.5}$$

where equation (4.3) is the normalization equation, and expression (4.5) is a set of n equations.

4.2 SOLUTION OF THE EXPRESSION BY CALCULUS

4.2.1 Solution for the Discrete Probability Distribution

Differentiating equation (4.1) with respect to p_i gives

$$d(-\frac{S}{K}) = \sum_{i=1}^{m} (\ell n\ p_i + 1)\ dp_i = 0 \tag{4.6}$$

Differentiating equations (4.3) and (4.5) with respect to p_i, keeping x_i and $< f_j(x) >$ constant gives

$$\sum_{i=1}^{m} dp_i = 0 \tag{4.7}$$

$$\sum_{i=1}^{m} f_j(x_i)\ dp_i = 0 \qquad j = 1, 2, \ldots, n \tag{4.8}$$

We multiply equation (4.7) by $(-\lambda_0 - 1)$ and expression (4.8) by $-\lambda_j$ $(j = 1, 2, \ldots, n)$, where λ_i $(i = 0, 1, 2, \ldots, n)$ are arbitrary functions (the Langrangian Multipliers).

$$(-\lambda_0 - 1) \sum_{i=1}^{m} dp_i = 0 \tag{4.9}$$

$$-\lambda_j \sum_{i=1}^{m} f_j(x_i)\, dp_i = 0 \qquad j = 1, 2, \ldots, n \tag{4.10}$$

All expressions in (4.10) are added, giving

$$\sum_{j=1}^{n} -\lambda_j \sum_{i=1}^{m} f_j(x_i) = 0 \tag{4.11}$$

Equations (4.6), (4.9), and (4.11) are added

$$\sum_{i=1}^{m} (\ln p_i + 1)\, dp_i + (-\lambda_0 - 1) \sum_{i=1}^{m} dp_i$$

$$- \sum_{j=1}^{m} \lambda_j \sum_{i=1}^{m} f_j(x_i) = 0$$

Collecting terms gives

$$\sum_{i=1}^{m} [\ln p_i - \lambda_0 - \sum_{j=1}^{n} \lambda_j f_j(x_i)]\, dp_i = 0 \tag{4.12}$$

Equation (4.12) must be satisfied regardless of the variation dp_i. Therefore, the quantity in the parentheses is equated to zero.

$$\ell n\, p_i - \lambda_0 - \sum_{j=1}^{n} \lambda_j f_j(x_i) = 0$$

or

$$\ell n\, p_i = \lambda_0 + \sum_{j=1}^{n} \lambda_j f_j(x_i)$$

Inversion gives

$$p_i = e^{\lambda_0 + \sum_{j=1}^{n} \lambda_j f_j(x_i)} \tag{4.13}$$

Substituting equation (4.13) in (4.3), gives

$$\sum_{i=1}^{m} p_i = \sum_{i=1}^{m} e^{\lambda_0 + \sum_{j=1}^{n} \lambda_j f_j(x_i)} = 1$$

so that

$$e^{\lambda_0} \sum_{i=1}^{m} e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)} = 1$$

This can be given the form

$$e^{-\lambda_0} = \sum_{i=1}^{m} e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)} \tag{4.14}$$

or

$$\lambda_0 = -\ln \sum_{i=1}^{m} e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)} \tag{4.15}$$

The x's may be determined in two ways. In the first method equation (4.15) is differentiated with respect to λ_k (k = 1, 2, . . ., n)

$$-\frac{\partial \lambda_0}{\partial \lambda_k} = \frac{\sum_{i=1}^{m} f_k(x_i)\, e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)}}{\sum_{i=1}^{m} e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)}} \tag{4.16}$$

But, from equation (4.14)

$$\sum_{i=1}^{m} e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)} = e^{-\lambda_0}$$

Substituting this into equation (4.16), gives

$$-\frac{\partial \lambda_0}{\partial \lambda_k} = \frac{-\sum_{i=1}^{m} f_k(x_i)\, e^{\sum_{j=1}^{m} \lambda_j f_j(x_i)}}{e^{-\lambda_0}}$$

or

$$\frac{\partial \lambda_0}{\partial \lambda_k} = -\sum_{i=1}^{m} f_k(x_i)\, e^{\lambda_0 + \sum_{j=1}^{n} \lambda_j f_j(x_i)}$$

But, from equation (4.13)

$$P_i = e^{\lambda_0 + \sum_{j=1}^{n} \lambda_j f_j(x_i)}$$

Substituting this into the right hand side of equation (4.17), gives

$$\frac{\partial \lambda_0}{\partial \lambda_k} = -\sum_{i=1}^{m} f_k(x_i)\, P_i$$

But, from equation (4.5)

$$< f_k > = \sum_{i}^{m} f_k(x_i)\, P_i$$

Thus,

$$= \frac{\partial \lambda_0}{\partial \lambda_k} = < f_k > \quad k = 1, 2, \ldots, n \tag{4.18}$$

Substituting this into equation (4.16), gives

$$< f_k > = \frac{\sum_{i=1}^{m} f_k(x_i)\, e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)}}{\sum_{i=1}^{m} e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)}} \quad k = 1, 2, \ldots n \tag{4.19}$$

Expression (4.19), contains n equations in n unknown, $(\lambda_1, \lambda_2, \ldots \lambda_n)$. The solution of these then may be used in equation (4.15) to obtain λ_0. Thus, all constants $(\lambda\ , \lambda_1, \ldots \lambda_n)$ in equation (4.13)

$$P_i = e^{\lambda_0 + \sum_{j=1}^{n} \lambda_j f_j(x_i)} \tag{4.13}$$

are known, and equation (4.13) represents the required probability distribution.

Proceeding with the second method of solving for the x's, we begin again with equation (4.15).

$$\lambda_0 = -\ln \sum_{i=1}^{m} e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)} \tag{4.15}$$

Substituting into equation (4.13), gives

$$P_i = e^{\left[-\ln\left(\sum_{i=1}^{m} e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)}\right) + \sum_{j=1}^{n} \lambda_j f_k(x_i)\right]} \tag{4.20}$$

This may be used in equation (4.5) to give

$$\langle f_k \rangle = \sum_{i=1}^{m} f_k(x_i)\, e^{\left[-\ln\left(\sum_{i=1}^{m} e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)}\right) + \sum_{j=1}^{n} \lambda_j f_j(x_i)\right]} \tag{4.21}$$

$$k = 1, 2, \ldots, n$$

Expression (4.21) contains n equations in n unknown, ($\lambda_1, \lambda_2, \ldots \lambda_n$). These may, in principle, be solved simultaneously, and, as before, equation (4.15) is used to obtain λ_0. So, again equation (4.13) is defined, and represents the required probability distribution.

4.2.2 Solution for the Continuous Probability Distribution

If the probabilities vary continuously, similar expressions can be obtained. Assume the values of x_i are uniformly spaced, giving

$$x_i = i \Delta x$$

where

$$i = i_{min}, i_{min} + 1, i_{min} + 2, \ldots\ldots, i_{max}$$

We add, and substruct $\ell n \Delta x$ on the right side of equation (4.15)

$$\lambda_0 = \ell n \Delta x - \ell n \sum_{i=1}^{m} e^{\sum_{j=1}^{n} \lambda_j f_j(x_i)} . \Delta x$$

We assume that Δx, is an infinitesimal.

$$\lambda_0 = \ell n \Delta x - \ell n \int_{x_{min}}^{x_{max}} e^{\sum_{j=1}^{n} \lambda_j f_j(x)} dx$$

Substituting this in equation (4.13), gives

$$P_i = e^{\left[\ell n \Delta x - \ell n \left(\int_{x_{min}}^{x_{max}} e^{\sum_{j=1}^{n} \lambda_j f_j(x)} dx \right) + \sum_{j=1}^{n} \lambda_j f_j(x) \right]}$$

$$\frac{P_i}{\Delta x} = e^{-\ell n \int_{x_{min}}^{x_{max}} e^{\sum_{j=1}^{n} \lambda_j f_j(x)} dx + \sum_{j=1}^{n} \lambda_j f_j(x)}$$

In the limit of small Δx, the continuous probability density function of x is

$$P(x) = e^{\lambda_0 + \sum_{j=1}^{n} \lambda_j f_j(x)} \tag{4.22}$$

where

$$\lambda_0 = -\ell n \int_{x_{min}}^{x_{max}} e^{\sum_{j=1}^{n} \lambda_j f_j(x)} dx \tag{4.22.a}$$

and all λ_j ($j = 1, 2, \ldots n$) satisfy relation (4.18) or relation (4.21), depending on which method is used.

Use of the first method leads to

$$\frac{\partial \lambda_0}{\partial \lambda_k} = \frac{\partial \left(-\ell n \int_{x_{min}}^{x_{max}} \exp\left(\sum_{j=1}^{n} \lambda_j f_j(x)\, dx\right)\right)}{\partial \lambda_k} = - \langle f_k(x) \rangle \tag{4.23}$$

$$k = 1, 2, \ldots n$$

and the second method results in,

$$\langle f_k(x) \rangle = \int_{x_{min}}^{x_{max}} f_k(x_i) \exp \left(\lambda_0 + \sum_{j=1} \lambda_j f_j(x) \right) dx \qquad (4.24)$$

$$k = 1, 2, \ldots n$$

Expression (4.23) or expression (4.24) consists of n equations in unknown λ_k ($k = 1, 2, \ldots n$). Solving them simultaneously gives the values for λ_k ($k = 1, 2, \ldots n$), and then λ_0 is obtained from equation (4.22).

Expression (4.23) or (4.24)(*) consists of n non-linear equations, and in general for such expressions more than one solution may exist. Each solution lies at a saddle point or a local extremum of the entropy function S. However, if it can be proven that the value of S obtained is a global maximum, then there will exist only one solution for expression (4.23) or expression (4.24).

4.3 PROOF THAT S IS A GLOBAL MAXIMUM(2)

If we consider two functions, S and G, defined as follows

$$S = - \sum p_i \ln p_i = \max \qquad G = - \sum g_i \ln g_i = \max$$

$$\sum p_i = 1 \qquad \sum g_i = 1$$

$$\sum f_r(x_i)\, p_i = \langle f_r(x) \rangle \qquad \sum f_r(x_i)\, g_i = \langle f_r(x) \rangle$$

$$p_i = \exp \left(\lambda_0 + \sum_j \lambda_j f_j(x_i) \right)$$

(*) also expressions (4.19) or (4.21), for discrete distributions.

Where the p_i's are defined in the last equation. The g_i represent any set of non-negative numbers which satisfy the equations of constraint. The problem now is to prove that S is greater than or equal to G for all possible g_i.

Consider the function

$$S - G = - \sum_i p_i \ln p_i + \sum g_i \ln g_i$$

We add and substract $\sum g_i \ln p_i$,

$$S - G = \sum (g_i - p_i) \ln p_i + \sum g_i \ln \left(\frac{g_i}{p_i}\right)$$

We know that $\ln p_i = \lambda_0 + \sum_r \lambda_r f_r(x_i)$. Substituting this for the first sum gives

$$S - G = \sum_i (g_i - p_i) \lambda_0 + \sum_i \sum_r \lambda_r (g_i - p_i) f_r(x_i)$$

$$+ \sum_i g_i \ln \left(\frac{g_i}{p_i}\right)$$

$$S - G = -\lambda_0 \sum p_i + \lambda_0 \sum_i g_i + \sum_r \lambda_r \left[\sum_i g_i f_r(x_i) \right.$$

$$\left. - \sum_i p_i f_r(x_i) \right] + \sum g_i \ln \left(\frac{g_i}{p_i}\right)$$

In view of the constaints, several terms cancel and therefore:

$$S - G = \sum_i g_i \ln \left(\frac{g_i}{p_i}\right) \tag{4.25}$$

Q_i is defined as

$$Q_i = g_i \ln g_i - g_i \ln p_i - g_i + p_i \tag{4.26}$$

Using $\sum p_i = 1$, and $\sum g_i = 1$, and (4.16) in equation (4.25), we get

$$S - G = \sum Q_i \tag{4.27}$$

We next differentiate equation (4.26) with respect to g_i twice, to give

$$\frac{\partial Q_i}{\partial g_i} = \ln g_i - \ln p_i$$

$$\frac{\partial^2 Q_i}{\partial g_i^2} = \frac{1}{g_i}$$

Therefore, the first derivative vanishes at $g_i = p_i$. Since f_i is always positive, the second derivative is always positive and therefore the point $g_i = p_i$ defines a minimum, not a maximum for Q_i. At $g_i = p_i$, Q_i is zero. Therefore the function Q_i is always non-negative and its minimum possible value is zero. Therefore (S - G) is non-negative for every choice of g_i and p_i, and zero for $g_i = p_i$. We can conclude that the solution for S is a global maximum, and there is only one solution for expression (4.23) or (4.24).

CHAPTER V

THE ALGORITHM FOR GENERATING THE MAXIMUM LOGARITHMIC ENTROPY DISTRIBUTION CONSTRAINED BY MOMENTS

5.1 INTRODUCTION

It may be recalled from Chapter IV, that usually the information available from estimating probabilities is in the form of averages of certain functions, more specifically not any type of functions, but rather the central moment functions. In this thesis the central moment functions are used as contraints to generate the distribution for two main reasons. We wish to be able to use the studies and the work done in the area of statistics and decision making, for example see Chapter II, and we wish to make a comparison between the new method, and the existing methods which use the moments as prior information.

Before proceeding to the algorithm, some mathematical relations are established to be used later.

5.1.1 The Relation Between the m^{th} Moment about the Origin and about the Expected Value

We define the following quantities

x is a random variable

$< x >$ is the expected value of x

By the binomial theorem,

$$(x - < x >)^m = x^m + (-1)\, m\, x^{m-1} < x > + (-1)^2 \frac{m(m-1)}{2!} x^{m-2} < x >^2$$

$$+ \ldots + \ldots + (-1)^m < x >^m$$

The right hand side can be rewritten as,

$$(x - \langle x \rangle)^m = \sum_{k=0}^{m} (-1)^k \frac{m!}{k!\,(m-k)!} x^{m-k} \langle x \rangle^k$$

Taking the expected value of both sides,

$$\langle (x - \langle x \rangle)^m \rangle = \sum_{k=0}^{m} (-1)^k \frac{m!}{k!\,(m-k)!} \langle x^{m-k} \rangle \langle x \rangle^k \qquad (5.1)$$

The term $\langle (x - \langle x \rangle)^m \rangle$ presents the m^{th} moment about the expected value, and the term $\langle x^{m-k} \rangle$ presents the $(m - k)$ th moment about the origin. A Fortran program (Appendix C) is used to calculate the moments about the origin from the moments about the expected value using equation (5.1).

5.1.2 The Relation Between the Moments of a Distribution and the Moments of its Transform

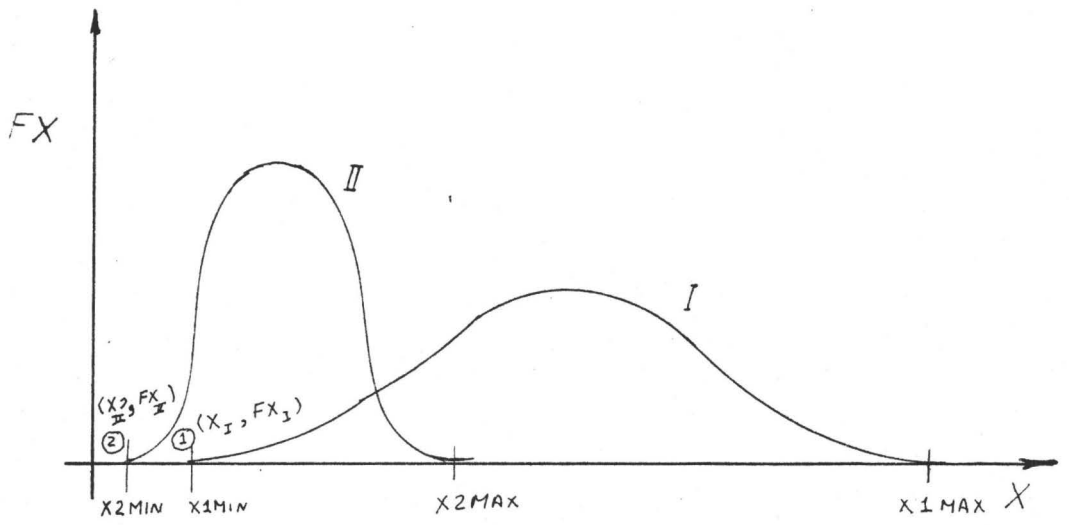


Figure 5.1 Probability distribution and its transform.

Assume curve I in Figure 5.1 represents a probability distribution, whose lower and higher bounds are X1MIN, and X1MAX, respectively. This curve is transformed to another domain, curve II in Figure 5.1, whose lower and higher bounds are X2MIN, and X2MAX, respectively. The relationship between the moments of curve I and curve II in Figure 5.1 will yield the relationship between the original and transformed moments.

It is obvious from Figure 5.1 that the relation between the transformed point 2, in curve II, and its original point 1, in curve I is

$$X_{II} = X2MIN + \frac{X_I - X1MIN}{S}$$

$$FX_{II} = S + FX_I$$

where

$$S = \frac{X1MAX - X1MIN}{X2MAX - X2MIN}$$

The first moment is a location factor, and therefore the first moments are related by

$$C^1_{II} = X2MIN + \frac{C^1_I - X1MIN}{S} \qquad (5.2)$$

This can be generalized for the ith moments if we define the following

C^i_I is the i^{th} moment for the original curve

C_{II}^{i} is the i^{th} moment for the transformed curve

By definition

$$C_{II}^{i} = \int_{X2MIN}^{X2MAX} (X_{II} - C_{II}^{1})^{i} \, FX_{II} \, dx$$

Using (5.2) gives

$$C_{II}^{i} = \int_{\frac{X1MIN}{S}}^{\frac{X1MAX}{S}} \left(X2MIN + \frac{X_I - X1MIN}{S} - X2MIN - \frac{C_I^1 - X1MIN}{S}\right)^{i} S \, FX_I \, dx$$

$$= \int_{X1MIN}^{X1MAX} \left(\frac{X_I - C_I^1}{S}\right)^{i} FX_I \, dx$$

$$= \frac{1}{S^{i}} \int_{X1MIN}^{X1MAX} (X_I - C_I^1)^{i} \, FX_I \, dx$$

By the definition of central moments,

$$C_{II}^{i} = C_I^i / S^i \qquad (5.3)$$

Appendix C contains a Fortran Program to calculate the transferred moments from the original moments.

5.1.3 The Relation Between the Maximum Logarithmic Entropy Distribution and Its Transform

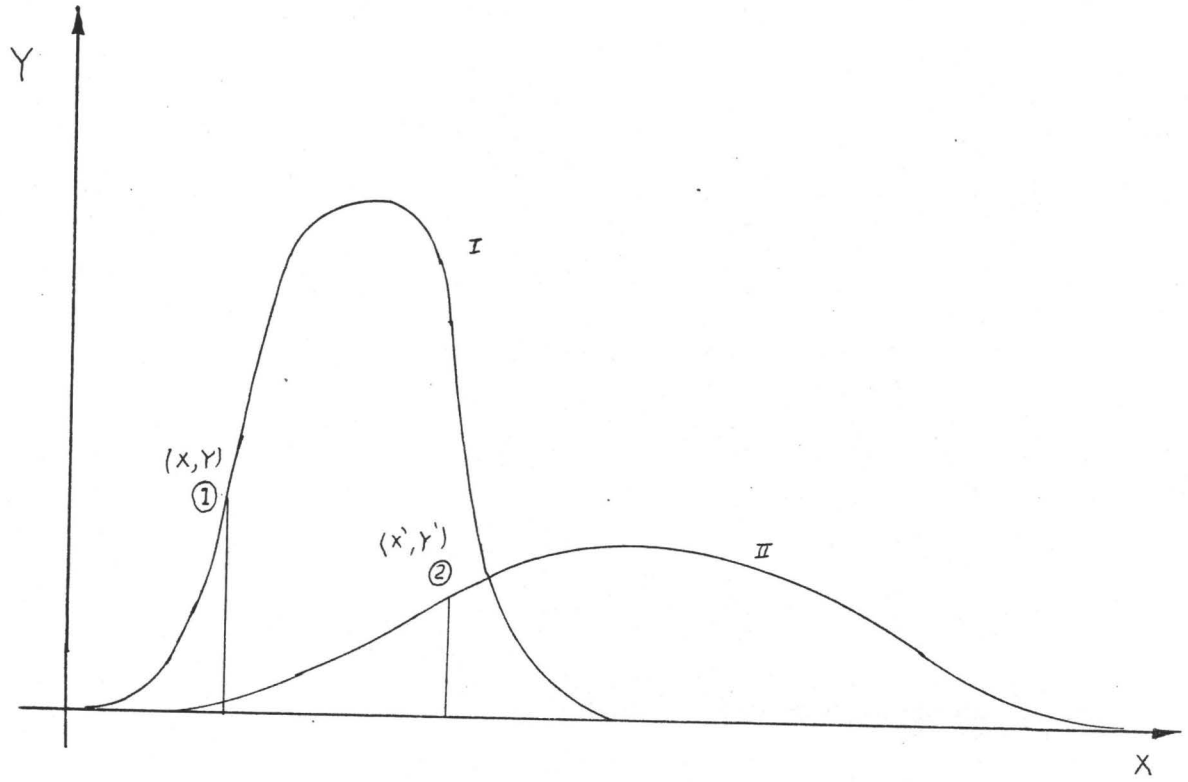


Figure 5.2 Probability distribution and its transform.

In Figure 5.2, if the probability distribution I is represented by the equation

$$y = \exp\left(\lambda_0 + \sum_{i=1}^{m} \lambda_i x^i\right) \tag{5.4}$$

and has lower and higher bounds XMIN and XMAX respectively. This distribution is transformed to the position II with a lower and higher bounds XMIN` and XMAX` respectively, and is represented by the equation

$$y = \exp\left(\lambda'_0 + \sum_{i=1}^{n} \lambda'_i x^i\right)$$

We wish to determine the relationship between the λ'_i's and λ_i's.

We first define

$$j = 1, 2, \ldots, n$$

$$S = \frac{XMAX' - XMIN'}{XMAX - XMIN}$$

$$A = \frac{S.XMIN - XMIN'}{S}$$

The relation between the transferred point (2) in curve II, and its original point (1) in curve I is

$$x = XMIN + \frac{X' - XMIN'}{S} = A + \frac{X'}{S}$$

$$Y = S\,Y'$$

Substituting in equation (5.4) gives

$$S\,y' = \exp\left(\lambda_0 + \sum_{i=1}^{n} \lambda_i \left(A + \frac{x'}{S}\right)^i\right)$$

or

$$y` = \exp\left[-\log S + \lambda_0 + \sum_{i=1}^{n}\left[\lambda_i \left(A^i + iA^{i-1}\frac{X`}{S}\right.\right.\right.$$

$$\left.\left.\left. + \frac{i(i-1)}{2!} A^{(i-2)}\left(\frac{X`}{S}\right)^2 + \ldots + \frac{i(i-1)(i-2)\ldots(i-k+1)}{k!}\right.\right.\right.$$

$$\left.\left.\left. A^{(i-k)}\left(\frac{X`}{S}\right)^k + \ldots + \left(\frac{X`}{S}\right)^i\right]\right]$$

Collecting terms, we get

$$y` = \exp\left(-\log S + \sum_{i=0}^{n} \lambda_i A^i + \sum_{j=1}^{n} \frac{X`^j}{S^j} \sum_{i=j}^{n}\right.$$

$$\left.\frac{i\,(i-1)\,(i-2\ldots(i-j+1)}{i!} A^{(i-j)}\lambda_j\right) \tag{5.5}$$

We could write equation (5.4) in the form

$$y` = \exp\left(\lambda`_0 + \sum_{j=1}^{n} \lambda`_j x^j\right) \tag{5.6}$$

Comparing (5.5) and (5.6), it follows that

$$\lambda_0 = -\ell n\, S + \sum_{i=0}^{n} \lambda_i A^i \tag{5.7}$$

$$\lambda'_1 = \frac{1}{S^j} \sum_{i=j}^{n} \frac{i\,(i-1)\,(i-2)\,\ldots\,(i-j+1)}{j!}\, A^{(i-j)}\, \lambda_j \qquad (5.8)$$

$$j = 1, 2, \ldots, n$$

In Appendix C, the last two equations are expressed in FORTRAN language.

5.2 SOLUTION FOR THE λ ' S

5.2.1 Expression Formulation

Recall from the previous chapter that the maximum entropy distribution is

$$p(x) = \exp\left(\lambda_0 + \sum_{j=1}^{n} \lambda_j x^j\right) \qquad (5.9)$$

where the λ 's satisfy the equations

$$\lambda_0 = -\ln \int_{x_{min}}^{x_{max}} \exp\left(\sum_{j=1}^{n} \lambda_j x^j\right) \qquad (5.10)$$

$$\frac{\partial \lambda_0}{\partial \lambda_k} = -C\,C_k \qquad k = 1, 2, \ldots, n \qquad (5.11)$$

Using the Simpson's rule multipliers for numerical integration to evaluate the integration in equation (5.10), we get

$$\lambda_0 = - \ln \sum_{i=1}^{m} [S_i \exp(\lambda_0 + \sum_{j=1}^{n} \lambda_0 x_i^j)] \tag{5.12}$$

where

S_i = the Simpson's rule multipliers

m = number of integration stations

Equation (5.12) is differentiated with respect to λ_j ($j = 1, 2, \ldots, n$), holding all other variables constant.

$$\frac{\partial \lambda_0}{\partial \lambda_j} = - \frac{\sum_{i=1}^{m} S_i x_i^j \exp\left(\sum_{j=1}^{n} \lambda_j x_i^j \right)}{\sum_{i=1}^{m} S_i \exp\left(\sum_{j=1}^{n} \lambda_j x_i^j \right)} \tag{5.13}$$

where $j = 1, 2, \ldots, n$

Substituting in equation (5.11), gives

$$C C_j - \frac{\sum_{i=1}^{m} S_i x_i^j \exp\left(\sum_{j=1}^{n} \lambda_j x_i^j \right)}{\sum_{i=1}^{m} S_i \exp\left(\sum_{j=1}^{n} \lambda_j x_i^j \right)} = R_j \tag{5.14}$$

$j = 1, 2, \ldots, n$

where R_j is a residual function equal to a very small number. Squaring and summary over all the j's, we get

$$R = \sum_{j=1}^{n} R_j^2 = \sum_{j=1}^{n} \left[C\,C_j - \frac{\sum_{i=1}^{m} S_i\, x_i^j \exp\left(\sum_{j=1}^{n} \lambda_j\, x_i^j\right)}{\sum_{i=1}^{m} S_i \exp\left(\sum_{j=1}^{n} x_i^j\right)} \right]^2 \qquad (5.15)$$

A solution exists for $R \leq \varepsilon$, where ε is a very small number.

The $C\,C_j$'s vary in value* and to obtain a solution that has the same relative allowed error we divide equation (5.14) by $C\,C_j$. Then equation (5.15) would be

$$R = \sum_{j=1}^{n} R_j^2 = \sum_{j=1}^{n} \left[1 - \frac{\sum_{i=1}^{m} S_i\, x_i^j \exp\left(\sum_{j=1}^{n} \lambda_j\, x_i^j\right)}{C\,C_j \sum_{i=1}^{m} S_i \exp\sum_{j=1}^{n} \lambda_j\, s_i^j} \right]^2 \qquad (5.16)$$

By using equation (5.16) as an optimization function and solving it by some appropriate nonlinear programming technique, a solution can be obtained either at $R \leq \varepsilon$ or $R_j \leq \varepsilon$, where ε is a very small number.

---

* This point can be illustrated by a numerical example. Assume the $C\,C_j$'s are .5, 1E-2, 5E-5, 3.5E-10, and the solution is required when $R \leq 1.E\text{-}12$, a solution could be obtained at $C\,C_j$'s values equal to .5000004, .0100003, .0000504, .0000007. Although the errors in the first and second moment are negligible, they are significant in the third and fourth moment.

5.2.2 Gradient Evaluation

The most successful optimization algorithm found for this problem requires the evaluation of the optimization function and the gradient vector at any given point. Each element of the gradient vector is the partial differential of the optimization function with respect to a variable.

$$g_i = \frac{\partial R}{\partial \lambda_i} \tag{5.17}$$

Equation (5.16) is differentiated with respect to λ_i.

$$g_i = \sum_{j=1}^{n} \left[\left(1 - \frac{\sum_{i=1}^{n} \eta_i^j}{C C_j \sum_{i=1}^{n} \eta_i^0} \right) \times \right. \tag{5.18}$$

$$\left. \frac{\sum_{i=1}^{m} \eta_i^0 \sum_{j=1}^{n} \eta_i^{2j} - \left(\sum_{j=1}^{n} \eta_i^j \right)^2}{C C_j \left(\sum_{i=1}^{m} \eta_i^0 \right)^2} \right]$$

where

$$\eta_i^k = S_i \, x_i^k \exp \left(\sum_{k=1}^{n} \lambda_k \, x_i^k \right)$$

5.3 DEFINITION OF DOMAINS

We have seen in Section 5.1 how we may set up an optimization function to define the λ's. We are interested in finding the maximum-logarithmic entropy distribution of a random variable x, given the numerical values of its n first moments, and the lower and the higher bounds of x. It can be seen in equation (5.16), that it is required to evaluate x^{2n}, where n is the number of given moment, and x varies from the lower bounds to the higher bounds. We thus must evaluate x_{max}^{2n}. If we assume, for illustration, that n = 6 and $x_{max} = 10^7$, then $x_{max}^{2n} = 10^{84}$. If we examine equation (5.16), we find that this value will be multiplied by some other possibly large values, so computer overflow is likely to occur.

To overcome overflow, x should be bounded by two members less than 1, so that x^{2n} will be less than 1 at all times. If the random variable x varies between x_{max} and x_{min}, we shall call the range between x_{max} and x_{min} the original domain. To overcome the overflow difficulty, we shall solve the problem at other lower and higher bounds, $\grave{x}_{max}$ and $\grave{x}_{min}$, where $\grave{x}_{max} \leq 1.$ and $\grave{x}_{min} < 1.$ We shall call the range between x_{max} and x_{min} the modified domain.

5.4 STARTING POINT ASSUMPTION

Most of the nonlinear programming techniques require a starting point to start the optimization algorithm. Theoretically, the final solution does not depend on the starting(*) point (in other words, with different starting points, there is only one final solution); but in practice, the selection of a bad starting point could lead to a solution with an excessive computer time or no solution at all, depending on the optimization-algorithm used. Usually

(*) This is not true in the case of a local optimum solution, but in the case discussed in this thesis there is only one global optimum solution.

a good starting point is a point near to the final solution. Below, we shall introduce four methods used to select a starting point, in order to provide an alternative if one method fails.

5.4.1 The Normal Assumption Starting Method

This method is suitable for small n. It is based on the well known fact that a normal distribution approximately represents many distributions. Thus a normal distribution for a start should often work well. If C_1, C_2, . . . , C_n are the first n central moments for a distribution, the best normal distribution that satisfies these moments is

$$y = \exp (\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \ldots + \lambda_n x^n)$$

where

$$\lambda_1 = \frac{C_1}{C_2}, \quad \lambda_2 = -\frac{1}{2C_2}, \quad \lambda_3 = \lambda_4 = \ldots = \lambda_n = \text{zero}$$

5.4.2 The Uniform Assumption Starting Method

This method is for small and large values of n. Some of the distributions like the J shape and the U shape cannot be approximated by a normal, and a uniform distribution would be preferable. All λ's values are zero except λ_0.

5.4.3 The (n + 1) Points Starting Method

This method is suitable for only large n. If C_1, C_2, . . . , C_n are the first n central moments for a distribution, and x_{min}, and x_{max} are the lower and higher bounds respectively for this distribution, it is required to

find an approximate probability distribution curve that has the form

$$f(x) = \exp\left(\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \ldots + \lambda_n x^n\right) \tag{5.19}$$

and satisfies the above information.

We assume (n + 1) points equally distributed between the lower and the higher bounds and select corresponding values for f(x) that satisfy the given moments. The moments are defined by

$$\int x\, f(x)\, dx = C_1$$

$$\int (x - C_1)^j\, f(x)\, dx = C_j \quad j = 2, 3, \ldots, n$$

$$\int f(x) = 1$$

The integrals may be approximated by Simpson's rule for (n + 1) points

$$\sum_{i=1}^{n+1} S_i\, x_i\, f_i = C_1 \tag{5.20}$$

$$\sum_{i=1}^{n+1} S_i\, (x_i - C_1)^j\, f_i = C_j \quad j = 2, 3, \ldots, n \tag{5.21}$$

$$\sum_{i=1}^{n+1} S_i\, f_i = 1 \tag{5.22}$$

where

S_i's the Simpson's rule multipliers

x_i's the assumed (n + 1) points

f_i's the corresponding density function values

Equations (5.20), (5.21), and (5.22) are (n + 1) linear equations in (n + 1) unknown (f_i's). Using Cramer's rule these values can be obtained. Substituting these values in equation (5.19), another (n + 1) linear equation can be formulated and solved to get the values of λ's.

5.4.4 Step By Step Starting Method

This method is suitable for only large n, and when all other methods fail to drive a solution. The method starts by obtaining the Maximum Logarithmic Entropy Distribution which satisfies the first two moments only. The resulting λ's, together with $\lambda_3 = 0$, are used as a starting point for a new Maximum Logarithmic Entropy Distribution which satisfies the first three central moments. This is repeated, increasing the number of moments and finding the corresponding Maximum Logarithmic Entropy Distribution, until the number of moments equal to n.

5.5 THE OPTIMIZATION TECHNIQUE

Recall from Section 5.2.1 that an appropriate nonlinear programming technique is required to solve equation (5.16) for the final solution. An approximate technique is one which finds a solution in the least amount of computation time. Only two techniques have been found by the author which are

capable of achieving the above--the Jacobson and Oksman method[20], and the new Fletcher method[21]. Since the computation time in Jacobson and Oksman method was found less than in the new Fletcher method, it was decided to use the Jacobson and Oksman technique in the maximization of the logarithmic entropy function algorithm.

5.6 THE MAIN ALGORITHM

The main algorithm is described below step by step, and a flow chart for the algorithm is shown in Figure 5.3.

1) Transfer to the modified domain.

   First define the modified domain, then calculate the transferred moment in this domain from the original domain, using equations (5.2) and (5.3).

2) Calculate the moment about the origin.

   Using equation (5.1), calculate the moment about the origin from the moment about the expected value.

3) Formulate the problem.

   Using equation (5.15) formulate the optimization function, and equation (5.18) for the gradient. Set a tolerance value ε, and assume a starting point using one of the methods discussed.

4) Solve the problem.

   By using any appropriate nonlinear programming technique, start optimizing, checking after each iteration the residual values. If $R_j < \varepsilon$ ($j = 1, 2, \ldots, n$), the current values of λ's are the solution. If for any reason, the nonlinear programming technique fails to get $R_j < \varepsilon$ ($j = 1, 2, \ldots, n$), select another starting point using an alternate method discussed in Section 5.4, and start optimizing

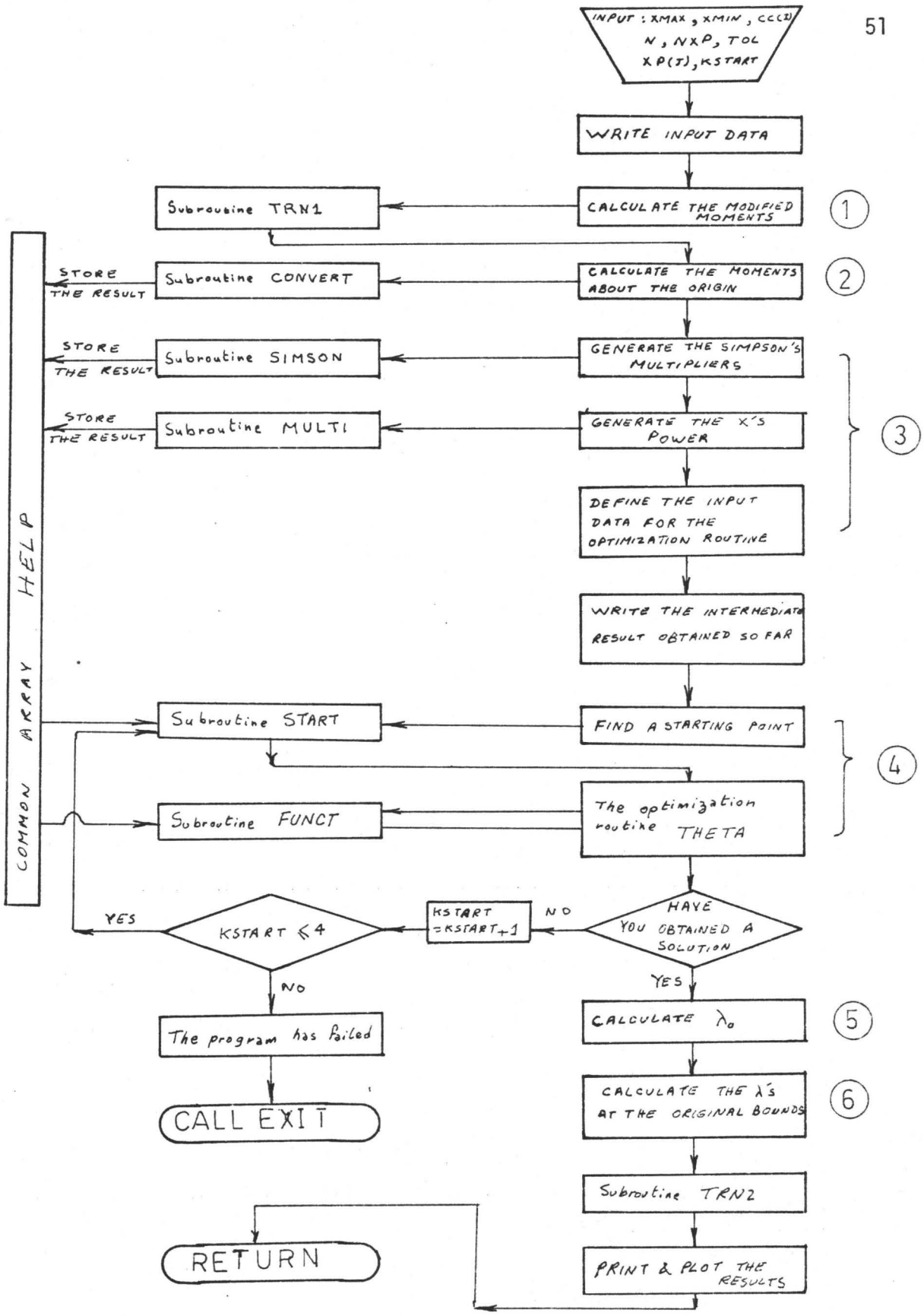


FIG. 5.3 M.L.E. algorithm flow chart

till $R_j < \varepsilon$ ($j = 1, 2, \ldots, n$).

5) Calculate λ_0.

Using equation (5.12), calculate λ_0.

6) Transfer back to the original domain.

Using equations (5.7) and (5.8), calculate all λ_j ($j = 0, 1, 2, \ldots, n$) at the original domain.

7) The solution.

With the values of the λ_j's obtained from the previous step, formulate the probability distribution expression, which is in the form

$$y = \exp (\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \ldots + \ldots + \lambda_n x^n)$$

where

x the independent variable

y the probability density function of x

λ_j constants, calculated from Step 7, $j = 0, 1, 2, \ldots, n$

CHAPTER VI

THE MAXIMUM LOGARITHMIC ENTROPY IN STATISTICAL MODELING FOR THEORETICAL POPULATION

6.1 GENERAL

Two methods are used to illustrate that the Maximum Logarithmic Entropy Distribution in an effective statistical model. The first method is by approximating the well known analytical distributions. The second method is by approximating the actual population and comparing these approximations with other existing methods. The second method will be considered in Chapter VII.

6.2 APPROXIMATING THE WELL KNOWN ANALYTICAL DISTRIBUTIONS

Most of the well known analytical distributions represent a population and are derived from actual populations of a specific type. An illustration of how the Maximum Logarithmic Entropy Distribution is an approximation to most of the analytical distributions, illustrates at the same time how the Maximum Logarithmic Entropy Distribution approximates the corresponding actual populations. Assuming that the analytical distributions are actual populations provides us with a variety of distributions amenable to the digital computer.

The first moments of an assumed actual distribution are calculated, and then the approximated Maximum-Logarithmic Entropy Distribution is generated from these moments. A comparison is then made between these two curves. For each analytical distribution, the first moments have been calculated from the following known relations

$$C_1 = \int_{x_{min}}^{x_{max}} x \; f(x)\, dx \tag{6.1}$$

$$C_i = \int_{x_{min}}^{x_{max}} (x - C_1)^i \; f(x)\, dx, \; i = 2, 3, \ldots, n$$

where

x independent variable

$f(x)$ probability density function

x_{min} the lower bound

x_{max} the upper bound

In the case where there is no definite value for the bounds, a reasonable value is taken so that the area beyond this value is negligible in comparison with the bounded area. The curve in the bounded area is then normalized. The approximated Maximum Logarithmic Entropy Distribution has been generated from these calculated moments with the same upper and lower bounds. Subroutine MEP in Appendix C has been used to generate the distribution, and the allowed relative error in the moment values for the solution has been taken equal to 10^{-6}.

Note that C_1 in the equation is the mean or expected value of a distribution, and is not a central moment whereas the C_i's are the central moments.

However, for convenience, the term central moments has been used to define all moments, including the expected value.

In some cases the first moment (the expected value) is presented as the only available information to generate the distribution. The result is usually an exponential distribution, as Tribus [3] demonstrated; but in some cases, where the mean value is in the midway between the lower and upper bounds, the generated distribution is found to be a uniform distribution. However this is a special case of the exponential.

Our knowledge of the distribution can logically be extended to the next higher moments in turn, and the distribution can be generated based on this knowledge. In each step a comparison is made between the assumed actual analytical distribution and its approximated Maximum Logarithmic Entropy Distribution by computing the percentage area they have in common. The two distributions are plotted together to show the deviation of the approximated curve from the actual one. The following analytical distributions have been surveyed, with various parameters-gamma, beta, Weibull, Rayleigh, exponential, Cauchy, and log-normal. The Weibull is presented in detail in the following section to give a visual illustration of how the accuracy varies with the number of moments.

6.3 EXAMPLE: THE WEIBULL DISTRIBUTION

The Weibull Distribution is represented by the equation

$$f(x) = \frac{\eta}{\sigma} \left(\frac{x}{\sigma}\right)^{\eta - 1} \exp\left[-\left(\frac{x}{\sigma}\right)^{\eta}\right]$$

for $\eta = 2$, $\sigma = 1$, and upper and lower bounds = 4.00 and 0.0. The method discussed in Section 6.1 has been applied for $n = 1, 2, \ldots, 5$, where n is the number of known moments. The results are summarized in the following pages.

WEIBULL DISTRIBUTION (W3)

$$f(x) = \begin{cases} \frac{\eta}{\sigma}\left(\frac{x}{\sigma}\right)^{\eta - 1} \exp\left[-\left(\frac{x}{\sigma}\right)^{\eta}\right], \ \eta = 2.0, \ \sigma = 1.0, \ x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 0.88623 0.21460 0.06274

0.149436 0.12793

Standardized moment measures($\sqrt{\beta_1}$, β_2): 0.63108 3.24484

Upper and lower bounds: 4.00 0.00

Type of curve: (bell shaped)

Table 6.1 Common area between Weibull Distribution (W3), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
1	69.96
2	94.66
3	96.52
4	97.76
5	98.22

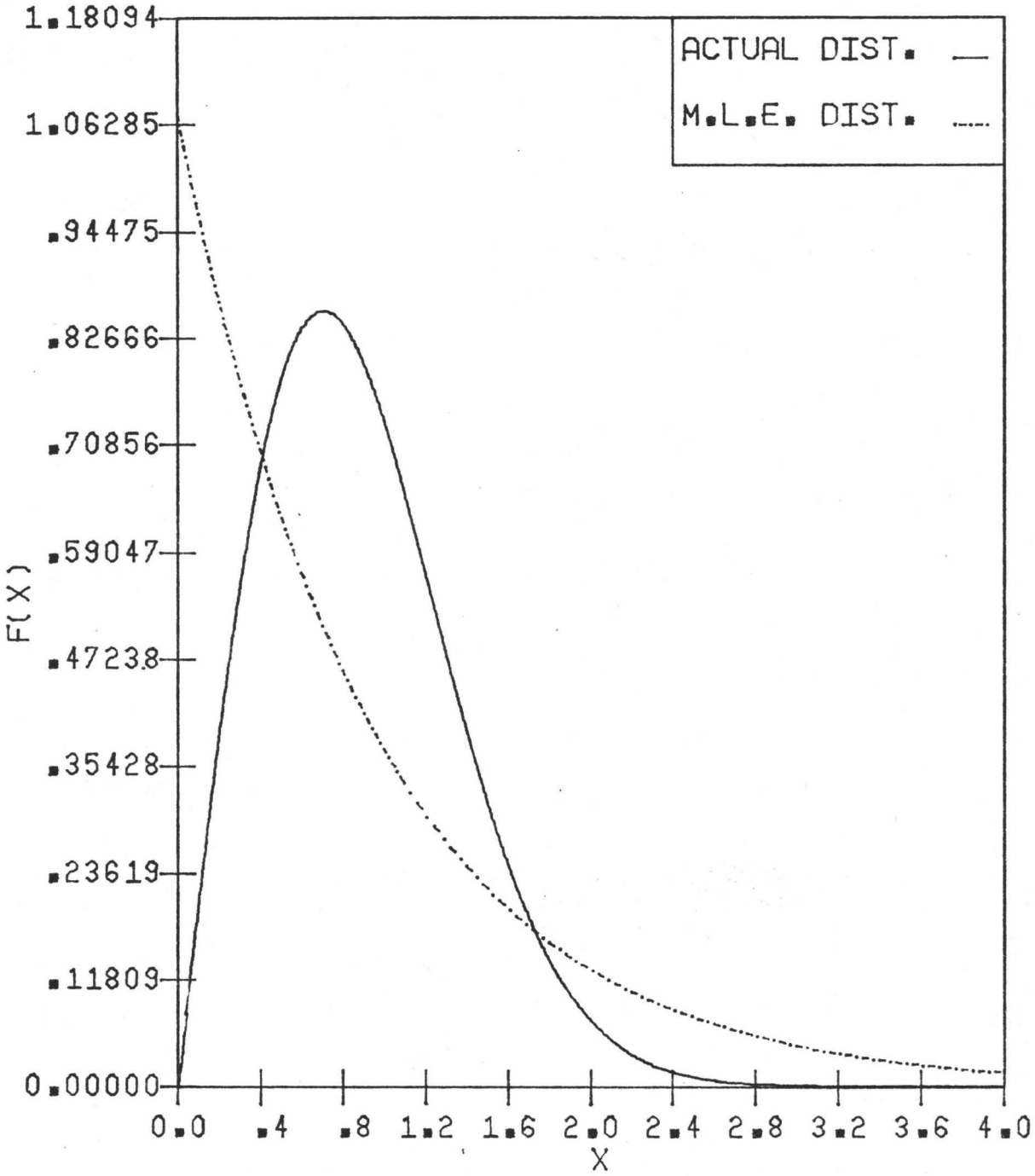


Figure 6.1 Approximating Weibull Distribution (W3), (η = 2.0, σ = 1.0), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first moment

Central Moment value: .88623

λ values (for M.L.E. Distribution): 0.071002 - 1.05799

Percentage area in common between the two distributions = 69.96

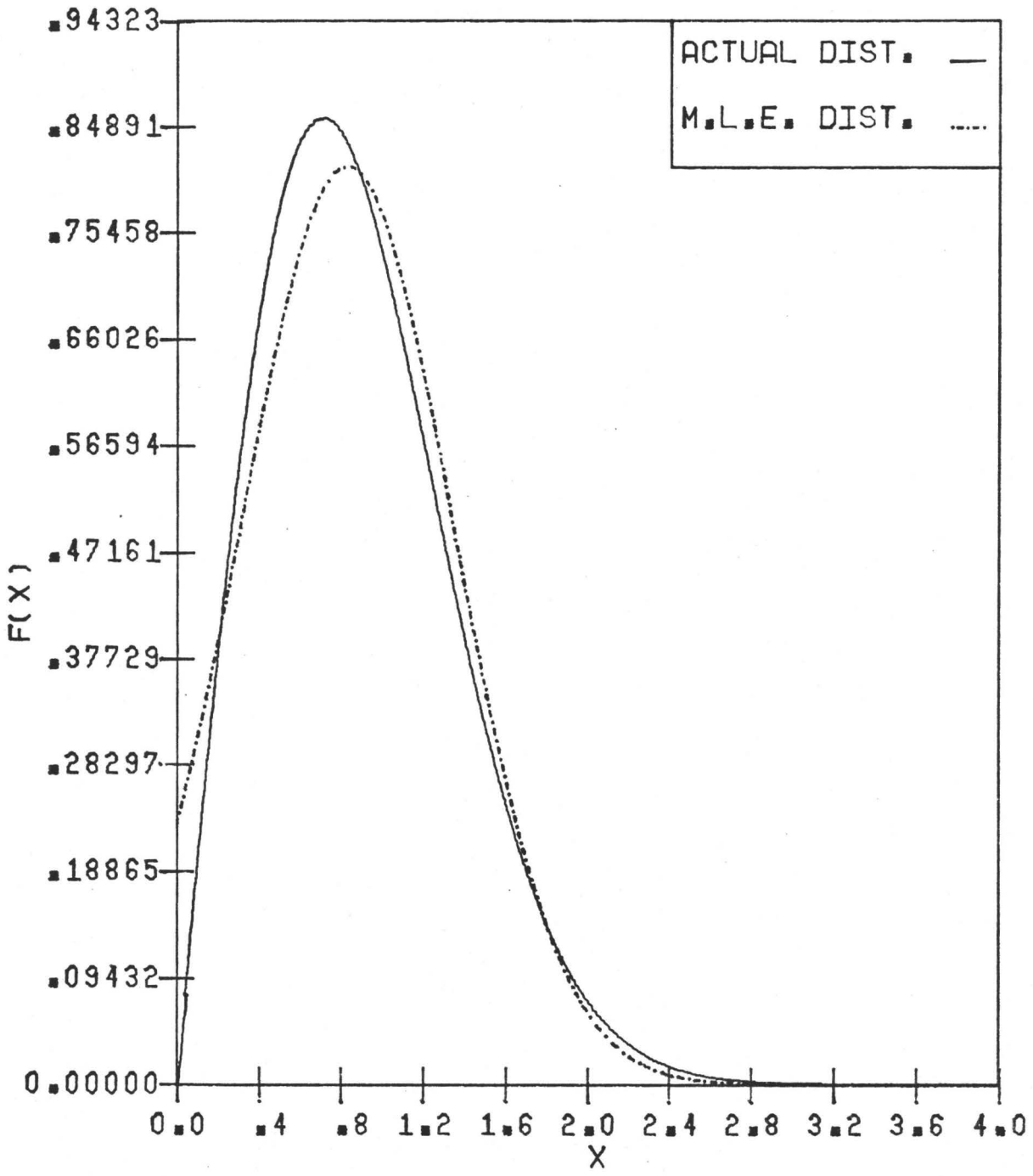


Figure 6.2 Approximating Weibull Distribution (W3), (η = 2.0, σ = 1.0), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first two moments

Central Moment values: 0.88623 0.21460

λ values (for M.L.E. Distribution): - 1.46238 + 3.05017 - 1.85157

Percentage area in common between the two distributions = 94.66

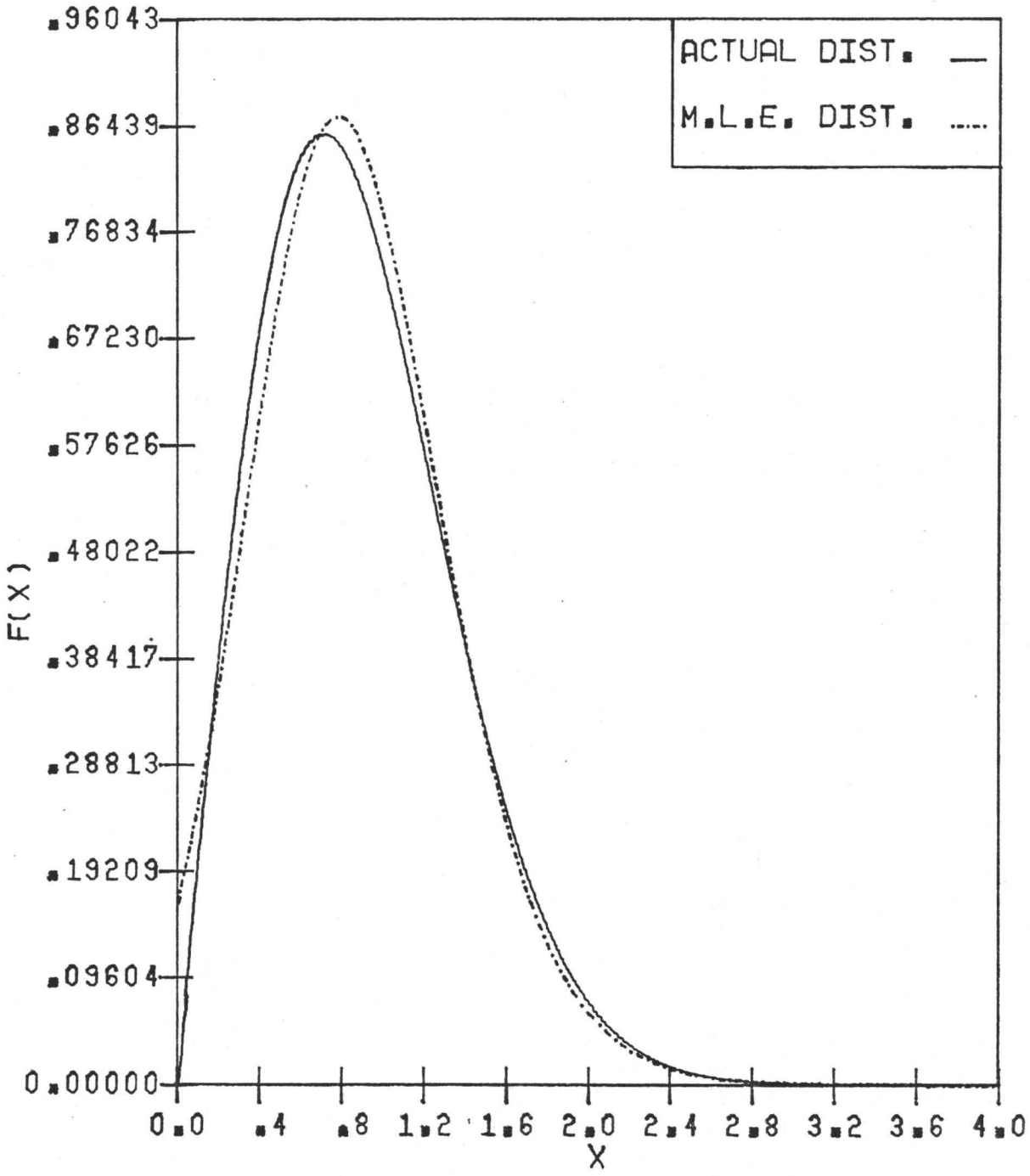


Figure 6.3 Approximating Weibull Distribution (W3), (η = 2.0, σ = 1.0), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first three moments

Central Moment values: .88623 .21460 .06274

λ values (for M.L.E. Distribution): - 1.81044 + 4.59126 - 3.51649 + .49278

Percentage area in common between the two distributions = 96.52

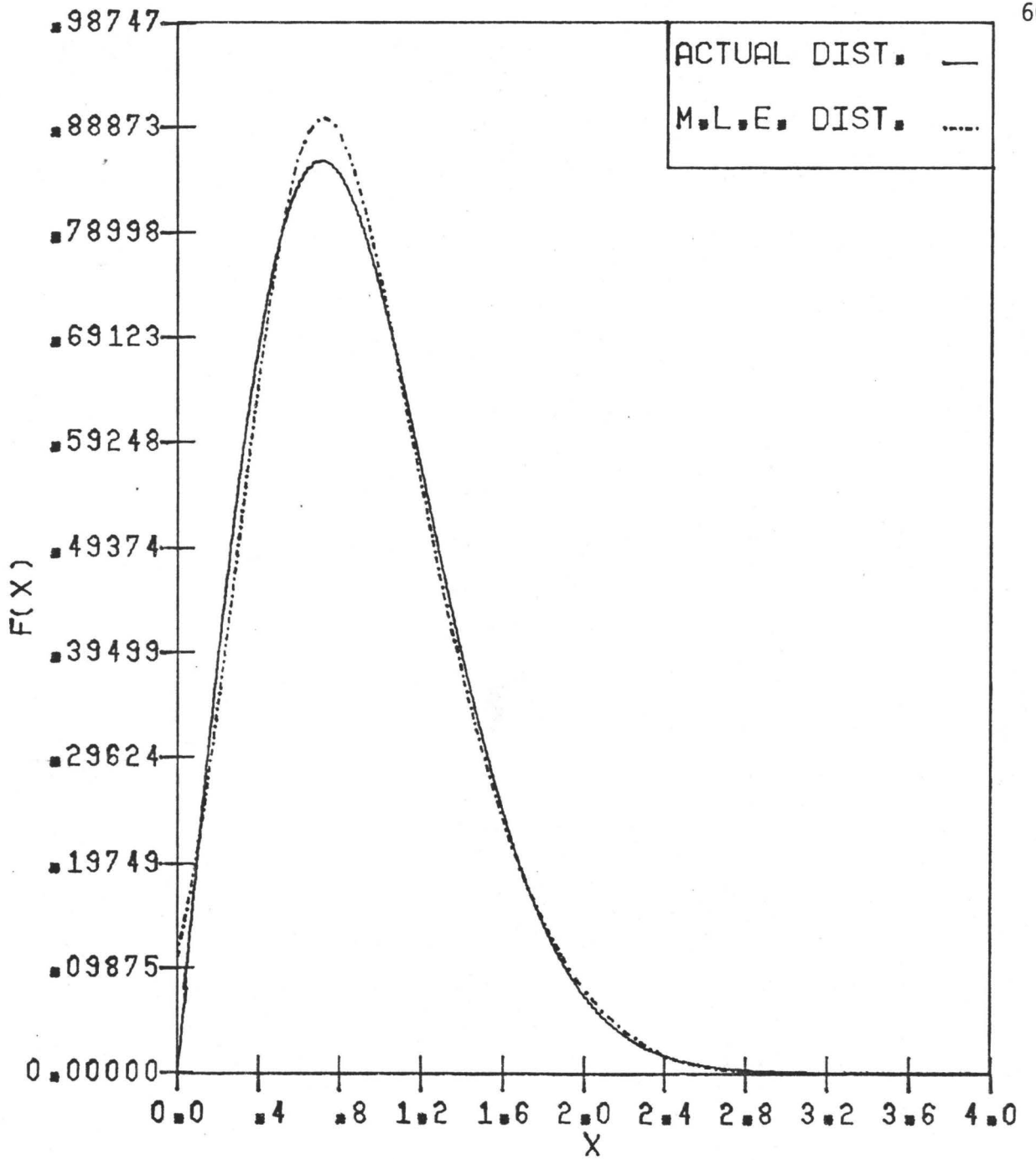


Figure 6.4 Approximating Weibull Distribution (W3), (η = 2.0, σ = 1.0), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first four moments

Central Moment values: .88623 .21460 .06274 .14944

λ values (for M.L.E. Distribution): - 2.20213 + 7.01450 - 7.57698 + 2.94650 - .47666

Percentage area in common between the two distributions = 97.76

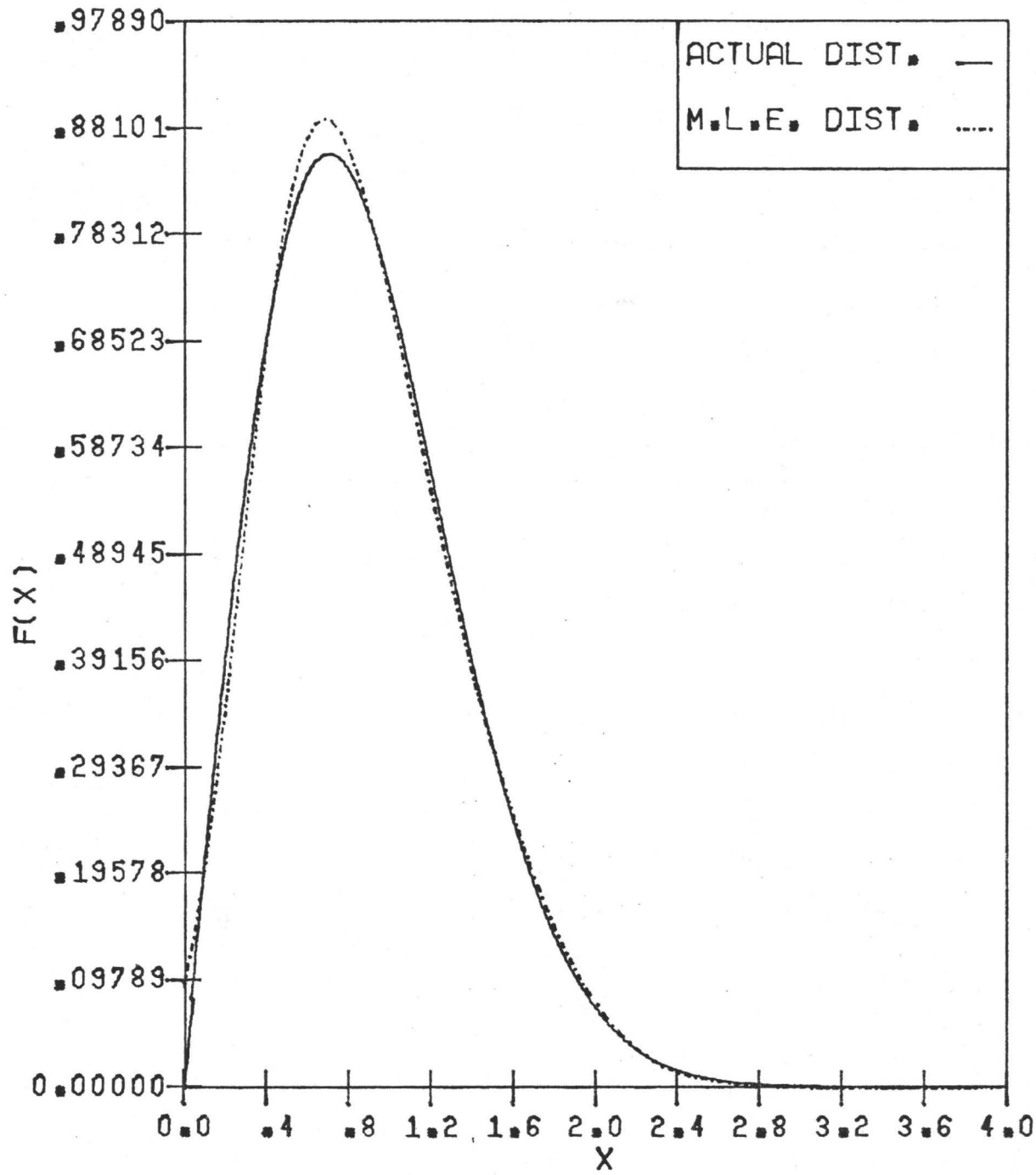


Figure 6.5 Approximating Weibull Distribution (W3), (η = 2.0, σ = 1.0), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .88623 .21460 .06274 .14944 .12793

λ values (for M.L.E. Distribution): - 2.39810 + 8.65600 - 11.51497 + 6.76357

Percentage area in common between the two distributions = 98.22

6.4 DISCUSSION OF THE WEIBULL EXAMPLE

The accuracy of the approximated curve, which is represented by the area in common between the Maximum Logarithmic Entropy Distribution and the assumed actual distribution, increases as the number of known moments increases. This would be anticipated since higher moments represent increasing knowledge about the independent variable (the Weibull distribution in this case). The above can be observed in Table 6.1.

For only one known moment (the expected value), the result is as expected, an exponential distribution; and for two known moments the result is a normal distribution (see reference (2)).

The accuracy of 98.22 per cent for five moments is a good approximation, and shows the Maximum Logarithmic Entropy as a reliable approximating method in statistical modeling.

6.5 GENERAL SURVEY OF THE APPROXIMATED ANALYTICAL DISTRIBUTIONS

The method discussed in Section 6.2 has been used for different types of analytical distributions, and different shape parameters (if there are any) for each distribution. For each curve the following procedure is used:

1) A symbol is given, to designate each curve. They are listed in Table A.1 (Appendix A).
2) The Maximum Logarithmic Entropy Distribution has been predicted for different number of known moments. The actual and its approximated Maximum Logarithmic Entropy Distribution are plotted together to show the deviation of the approximate curve from the original one. The

plots are shown in Appendix A (Figure A.1 to Figure A.34). Note that for each curve only one comparison figure between the original and the approximated curves is plotted, unless a criteria would be demonstrated.

If we wish to know how to choose the Maximum Logarithmic Entropy Distribution approximation is to any analytical distribution stated in Table A.1, the following steps should be followed.

1) Find the symbol for the required distribution in Table A.1.
2) Find this symbol in (β_1, β_2) plane (Figure 6.7) and note beside it the symbol suggesting the shape of the curve and the number for the percentage area in common between the analytical curve and its approximation Maximum-Logarithmic Entropy Distribution.
3) Table 6.2 also gives the figure number for the illustration of the two curves, and the table member for the comparison between the two curves for different known first moments.

The variation of the parameters involved in the expression for most analytical distributions, leads to different shapes for the same general distribution. However, most of them are similar. To illustrate most of the known distributions in the restricted space of Figure 6.7 (β_1, β_2 plane), different specific values of these parameters have been chosen to represent the different shapes for any distribution in a minimum number of curves. As an example, the Weibull distribution can take an infinite number of shapes, most of them are similar, and are included in the four following categories:

1) the skewed bell-shaped type
2) the symmetric bell-shaped type
3) the exponential type

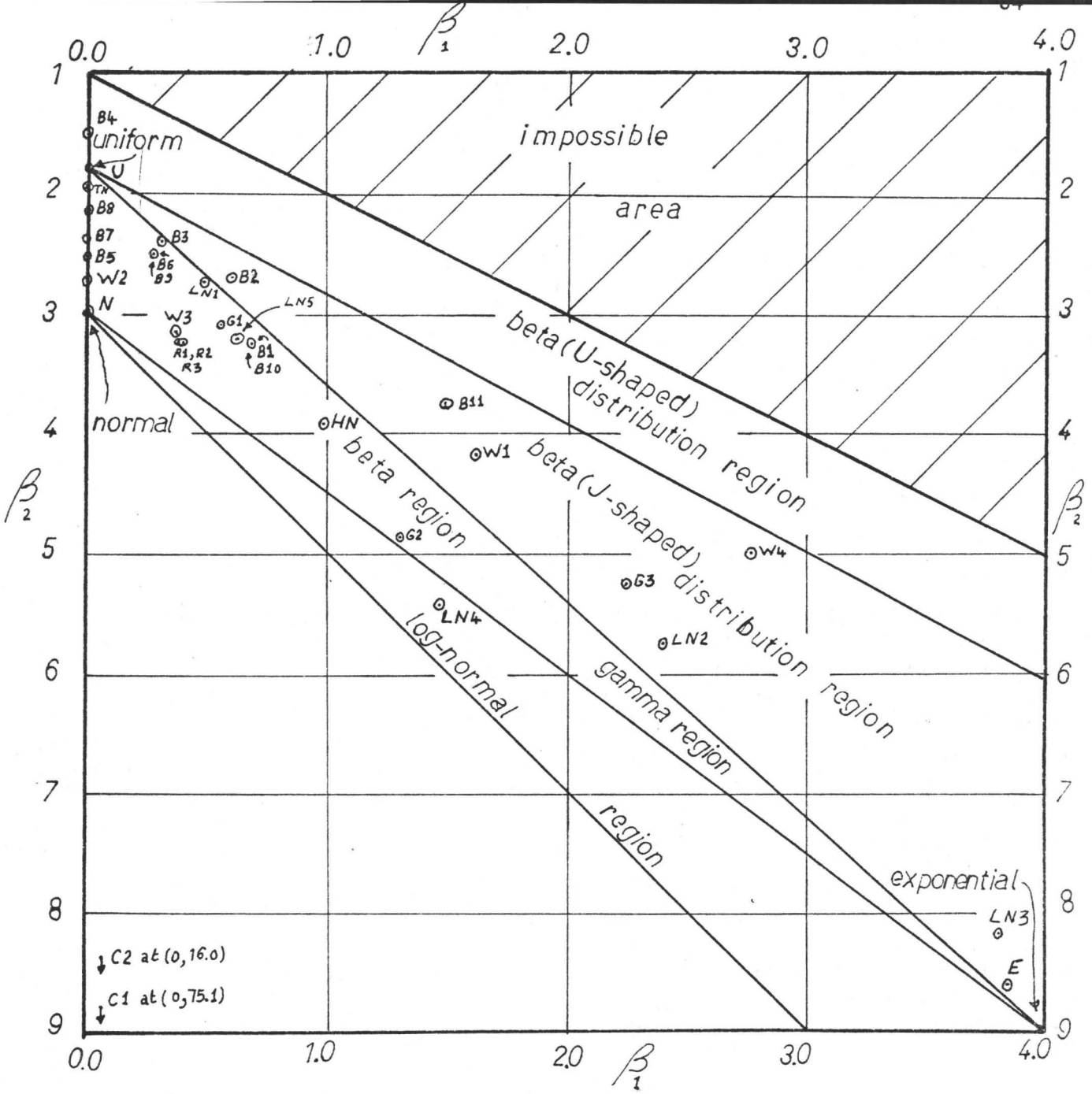


Figure 6.6 Points in (β_1, β_2) plane for various analytical distributions, each point represents a distribution and its approximate Maximum Logarithmic Entropy distribution. For the meaning of the symbols (near each point) see Table A.1. The points are plotted in the suggested regions for various distributions given by Professor E. S. Pearson, University College, London[4].

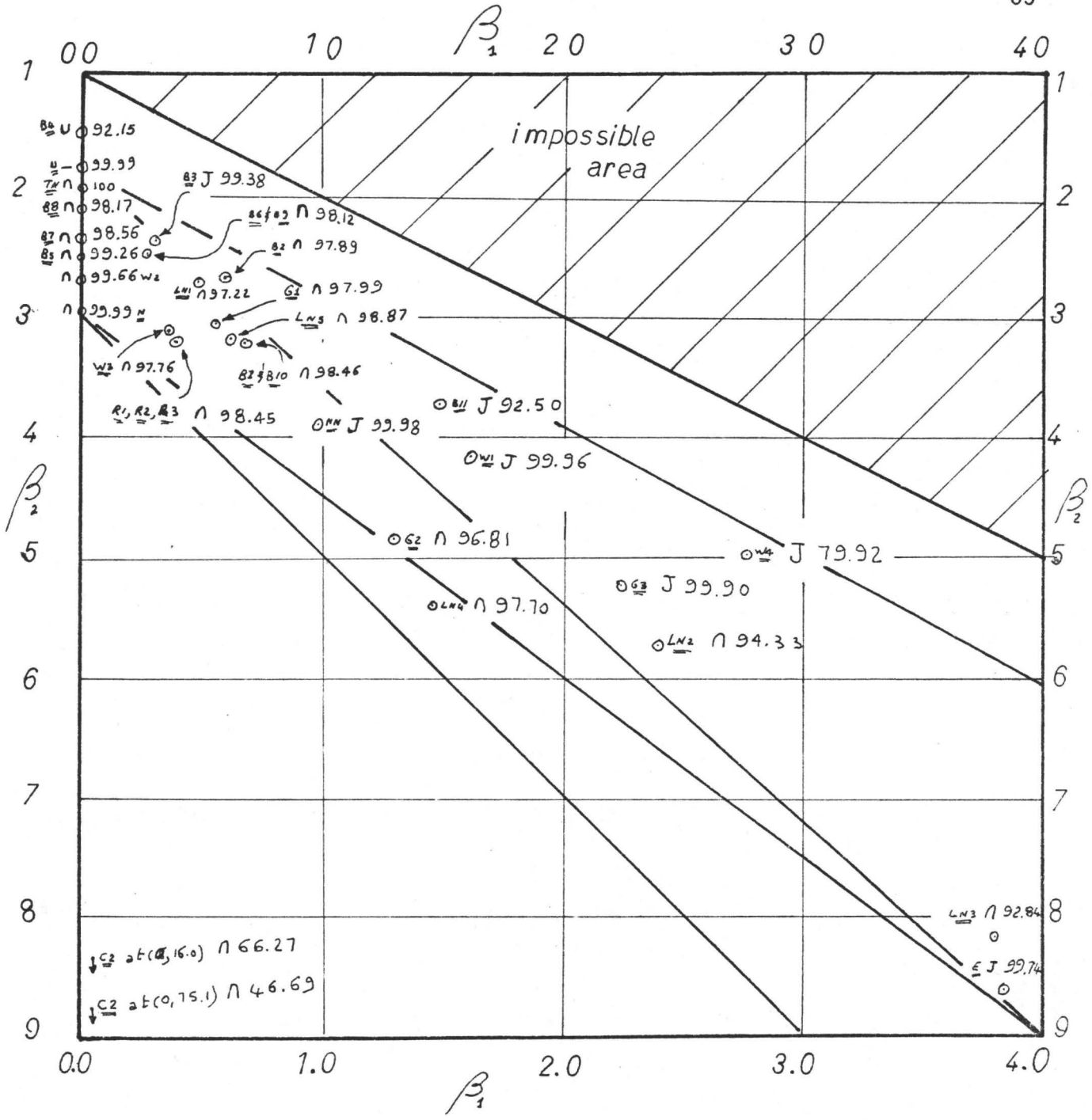


Figure 6.7 Points in (β_1, β_2) plane for various analytical distributions approximated by the Maximum Logarithmic Entropy Method. Symbols J, Λ, V denote J-shaped, bell type shaped, U-shaped, and uniform shaped distributions, followed by a number which is the percentage area in common between the distribution and its approximate Maximum Logarithmic Entropy distribution, for the first five moments.

4) the J-shaped type

So according to the above condition only four shape parameter (η) of the Weibull distribution have been chosen to present the distribution in Figure 6.7

6.6 DISCUSSION OF THE GENERAL SURVEY

The results illustrated in this chapter and in Appendix A, shows the flexibility of the Maximum Logarithmic Entropy Distribution to represent almost any kind of population. In cases of bell-shaped distributions, the area in common between the analytical distribution, and the Maximum Logarithmic Entropy Distribution having the same moments as the analytical one (or its approximate), varies between 96.81 per cent (in the case of the gamma distribution) and 100.00 per cent (in the case of the normal distribution), indicating that the Maximum Entropy Distribution is close to the assumed actual distribution. However the Cauchy Distribution is an exception to the above figures (see Figures A.27, A.28, and Tables A28, A29. It is a symmetric long-tailed distribution, and very rarely occurs in practice. Investigating the various distributions in Figure 6.7, we can see that all distributions concentrate in the area where $0.0 < \beta_1 < 4.0$, and $1.0 < \beta_2 < 9.0$, which we can consider the practical region for the probability distributions, and all the results inside this area are acceptable and reasonable, but as we move outside this region, the accuracy starts to decline, and the further we go the less accuracy we obtain. This is quite observable in the Cauchy distributions, the accuracy is better in the distribution near to the above region (C2) than the far one (C1).

In cases of J-shaped distributions the situation is not as good as in the bell-shaped distributions. The Maximum Logarithmic Entropy Distributions have an area correspondence with the actual ones from 100.00 per cent (in the case of the exponential distribution (E)) to 79.92 (in the case of the Weibull

distribution (W1)). In the case of the U shaped curve the deviation is reasonable, the percentage area in common between the two curves is 92.15 percentage.

In most of the cases the accuracy increases as the number of known moments increases. This would be anticipated since higher moments represent increasing knowledge about the independent variable. But, the error involved in finding a solution (error in Lagrangian Multipliers λ's) for higher moments is greater than the error in a solution for a lesser number of moments. In addition to the error obtained from calculating the distribution for higher moments is less than the error obtained from calculating the distribution for lower moments. In other words, if an error has occurred in the independent variable x, the expected error from calculating the dependent variable y for n moments is more than the expected error from calculating the dependent variable y for (n + 1) moments. This can be illustrated for n = 1

$$y_1 = \exp (\lambda_0 + \lambda_1 x)$$

$$y_2 = \exp (\lambda_0 + \lambda_1 x + \lambda_2 x^2)$$

Obviously the error in y_2 is more than the error in y_1, for a given error in x. Thus, although increasing knowledge about an independent variable helps in predicting its distribution more accurately, on the other hand, the error in the prediction process increases as our knowledge increases. So in each step of increasing knowledge there is a gain in accuracy counteracted by a loss also due to numerical error. If at a certain stage of knowledge the accuracy achieved its maximum, we can expect a decline in the accuracy in the next step of increasing knowledge. This is noticeable in the gamma distribution [G3,

Table (A.4)], where an accuracy of 100 per cent is achieved for the first two moment, but there is a decline as the number of moments increases. It can also be observed in the following distributions; Beta (B.7), exponential (E), Weibull (W1), uniform (U), normal (N), truncated normal (TN), half normal (HN). The amount of loss in accuracy is, however, not significant.

CHAPTER VII

STATISTICAL MODELING ON ACTUAL DATA AND COMPARISON WITH OTHER METHODS

7.1 GENERAL

This chapter will make a comparison between the Maximum Logarithmic Entropy Method and the Johnson and Pearson methods using problems in the references. The existing solutions are compared with solutions by the new method.

7.2 THE COEFFICIENT OF FRICTION PROBLEM

7.2.1 Comparison with Johnson Method

This problem is given by Hahn and Shapiro (reference (1), page 219). Measurements of the coefficient of friction for a metal were obtained on 250 samples. The actual and the predicted values by Johnson method are summarized in the first three columns in Table 7.1.

The following moments values were calculated from Table 7.1.

$$C_1 = 3.448 \times 10^{-2}$$

$$C_2 = 9.238 \times 10^{-5}$$

$$C_3 = 4.860 \times 10^{-7}$$

$$C_4 = 2.742 \times 10^{-8}$$

The upper and lower bounds are assumed from the given data

XMIN = .010

XMAX = .065

Table 7.1 Comparison of predicted per cent observations by frequency classes for Johnson's Distribution Fit to Coefficient of Friction for 250 samples versus predicted per cent observation by the Maximum Logarithmic Entropy Method.

Coefficient of Friction	Actual per cent of Observation	Predicted per cent of Observations	
		Johnson Method	Maximum Logarithmic Entropy Method
Less that 0.0150	0.4	0.7	.57
0.015 to 0.0199	3.6	3.3	3.51
0.020 to 0.0249	12.0	10.8	10.96
0.025 to 0.0299	17.6	20.0	19.51
0.030 to 0.0349	23.2	22.5	22.22
0.035 to 0.0399	18.0	17.9	18.06
0.040 to 0.0449	11.6	11.5	11.64
0.045 to 0.0499	6.8	6.4	6.57
0.050 to 0.0549	3.6	3.5	3.57
0.055 to 0.0599	1.6	1.7	2.05
0.060 or more	1.6	1.7	1.36

The Maximum Logarithmic Entropy Method is applied to the above data, and the results are shown in the fourth column in Table 7.2. The mathematical model was found to be

$$y = \exp (\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3 + \lambda_4 x^4) \qquad (7.1)$$

where

$\lambda_0 = -8.207$

$\lambda_1 = 921.901$

$\lambda_2 = -22624.2$

$\lambda_3 = 182945.1$

$\lambda_4 = -308036.5$

The comparison is made between the two methods by calculating the absolute error at each point for the two methods, shown in the second and fourth columns in Table 7.2 and plotted on Figure 7.2.

It is quite clear that the M.L.E. is a better approximation than the Johnson method except in the tail end.

In general, to judge which method is better the goodness of fit test is conducted. The result is shown in the third and firfth columns of Table 6.2. The total χ^2 contribution in the M.L.E. (0.53) is less than the total χ^2 contribution in the Johnson method (.64). The result is also plotted in Figure 7.3.

The Maximum Logarithmic Entropy Method is thus better than the Johnson method for this example.

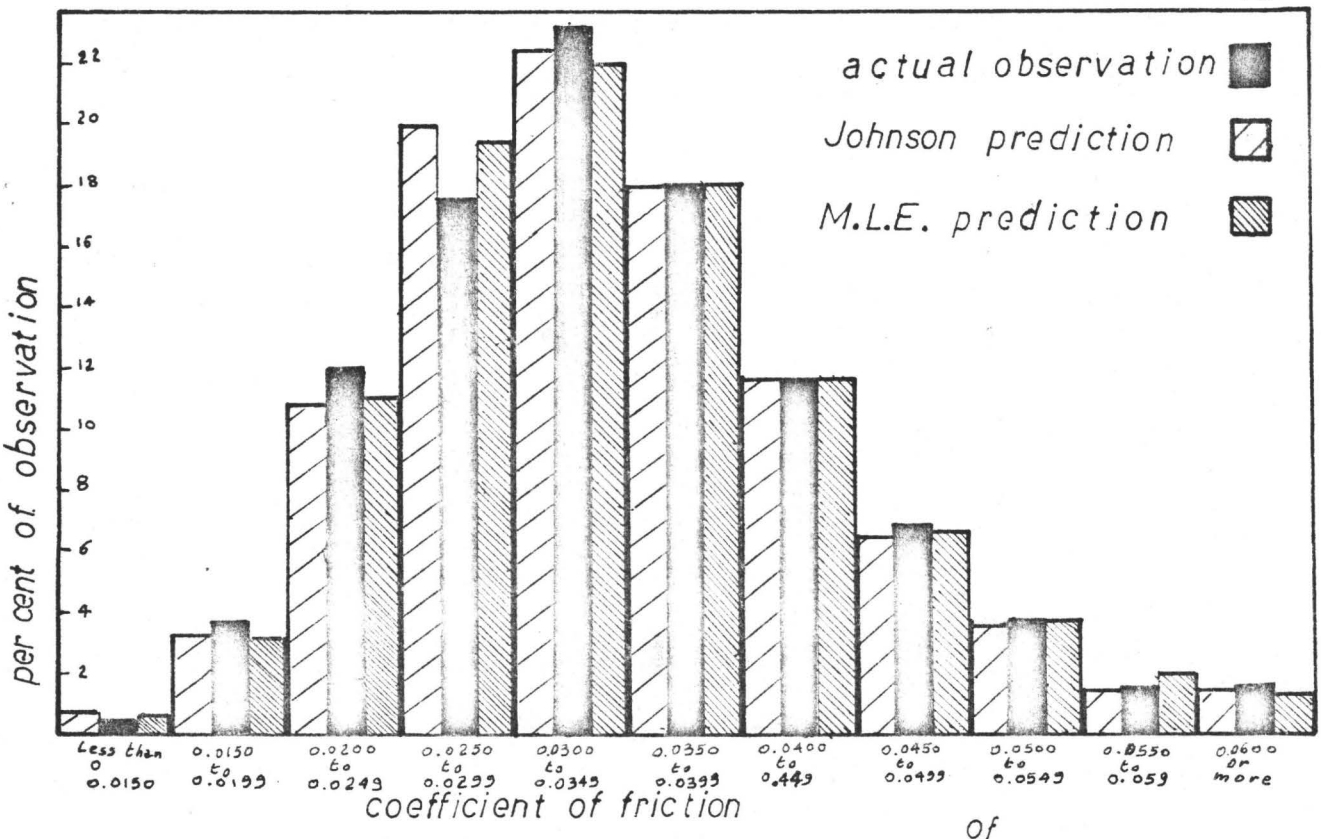


Fig.7.1 Actual versus Johnson and M.L.E. predicted per cent of coefficient of friction data

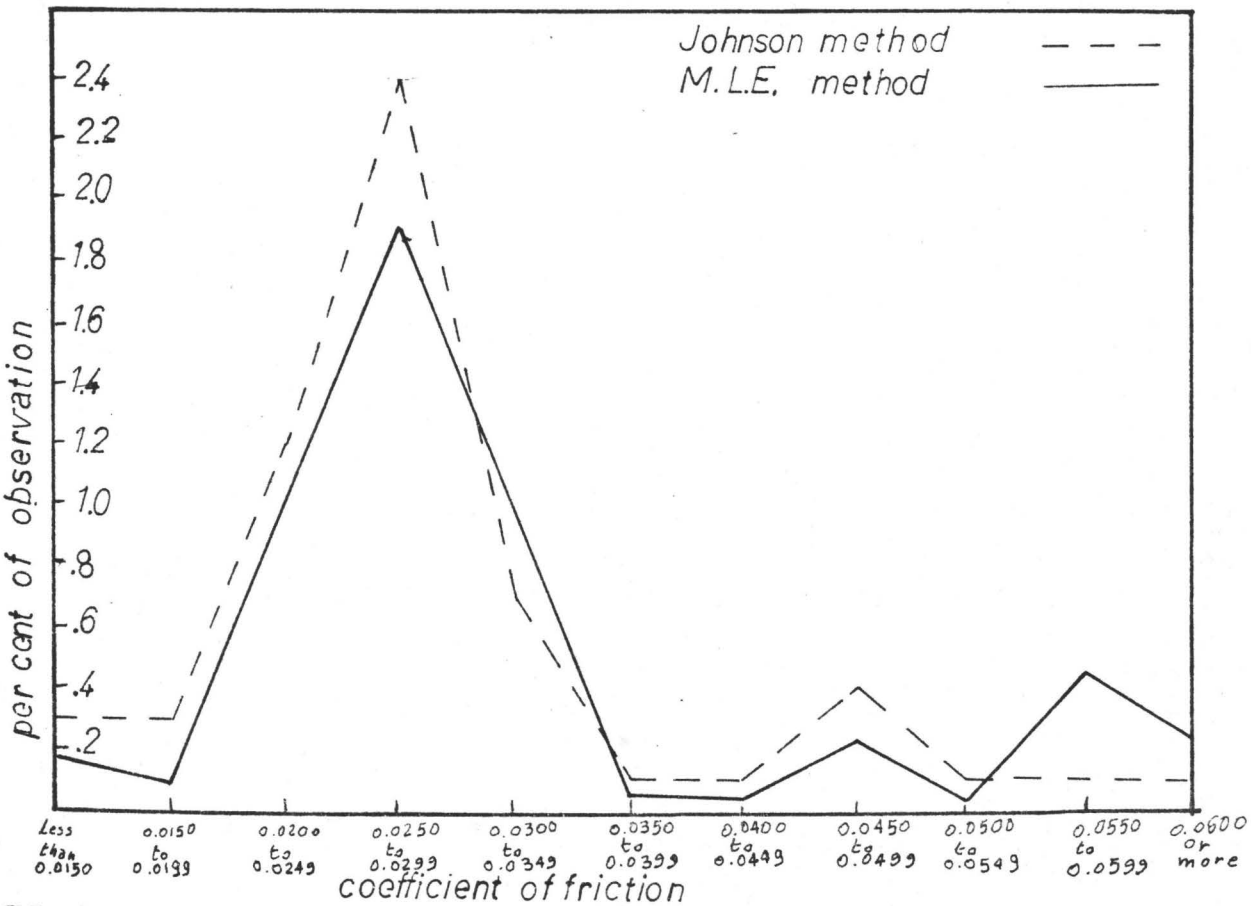


Fig.7.2 Johnson versus M.L.E. absolute errors in coefficient of friction problem

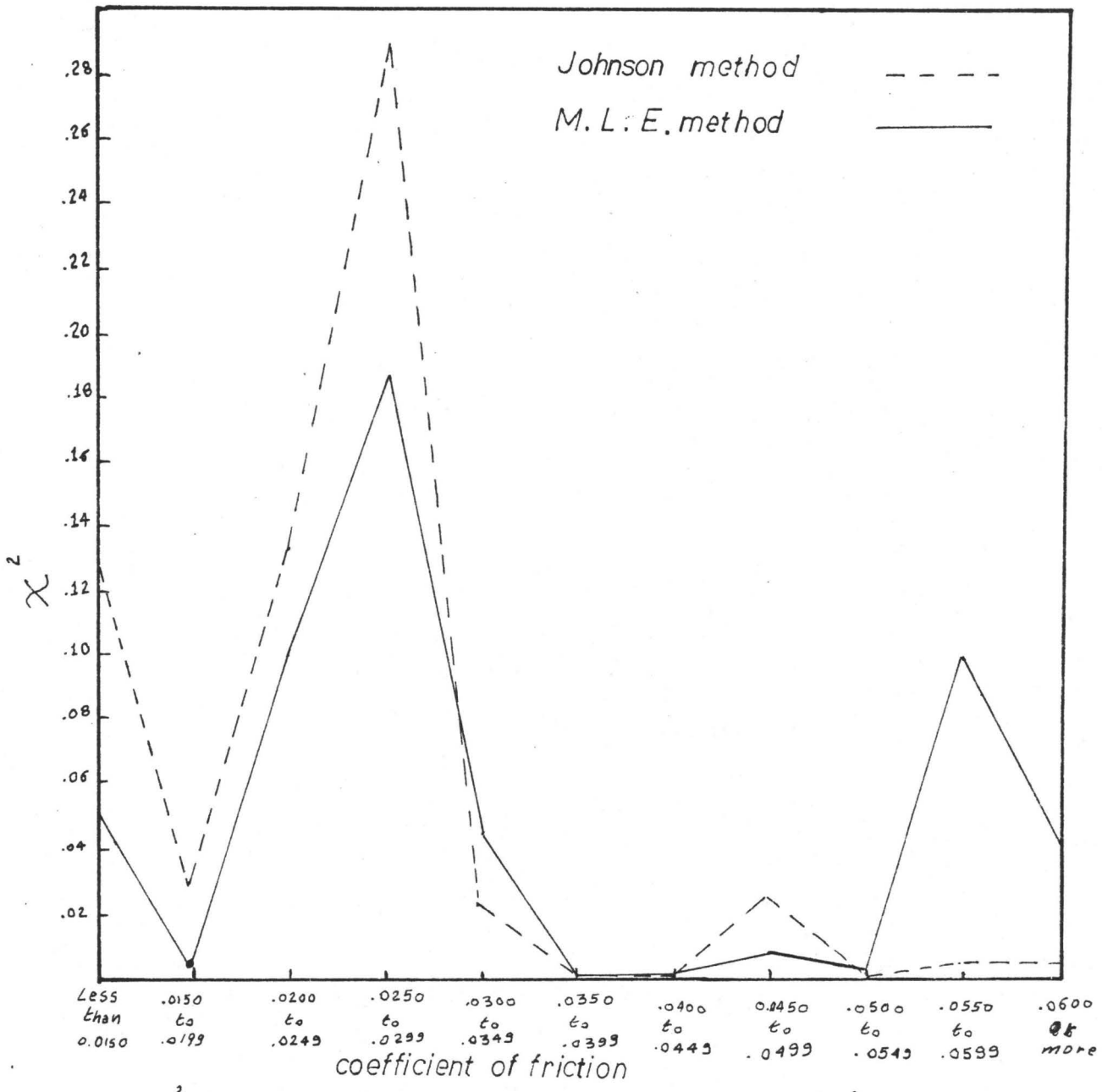


Fig. 73 χ^2 contribution in Johnson method versus χ^2 contribution in M.L.E. method for coefficient of friction problem

Table 7.2 Comparison of absolute errors and χ^2 contribution in Johnson Method versus absolute errors and χ^2 contribution in M.L.E. Method.

Actual per cent of Observation	Johnson Method		M.L.E. Method	
	Absolute error	χ^2 contribution	Absolute error	χ^2 contribution
0.4	0.3	.1286	.017	.0508
3.6	0.3	.0273	.09	.0023
12.0	1.2	.1333	1.04	.0998
17.6	2.4	.2880	1.90	.1867
23.2	0.7	.0218	0.98	.0430
18.0	0.1	.0006	.07	.0002
11.6	0.1	.0009	.04	.0001
6.8	0.4	.0250	.23	.0081
3.6	0.1	.0029	.02	.0002
1.6	0.1	.0059	.45	.1004
1.6	0.1	.0059	.24	.0427
		total χ^2 = .64		total χ^2 = .53

7.2.2 Comparison with Pearson Method

The above example is also solved by using Pearson's method. The cummulative percentages prediced by Pearson and Johnson were calculated at five points and are shown in the second, third and fourth column of Table 7.3. The corresponding cummulative percentages predicted by the Maximum Logarithmic Entropy Method are computed using equation (7.1), and are shown in the fifth column of Table 7.3.

Table 7.3 Comparison of Cummulative percentages from actual data, Pearson, Johnson, and Maximum Logarithmic Entropy Approximations for coefficient of friction data.

Variable X	Cummulative Percentages from actual data	Corresponding Cummulative Percentage predicted by		
		Pearson Method	Johnson Method	Maximum Logarithmic Entropy Method
.021	6.4	5.0	5.4	5.88
.023	11.2	10.0	9.3	9.91
.033	47.6	50.0	48.5	47.96
.052	94.8	95.0	94.8	94.61
.064	99.2	99.0	99.0	99.75

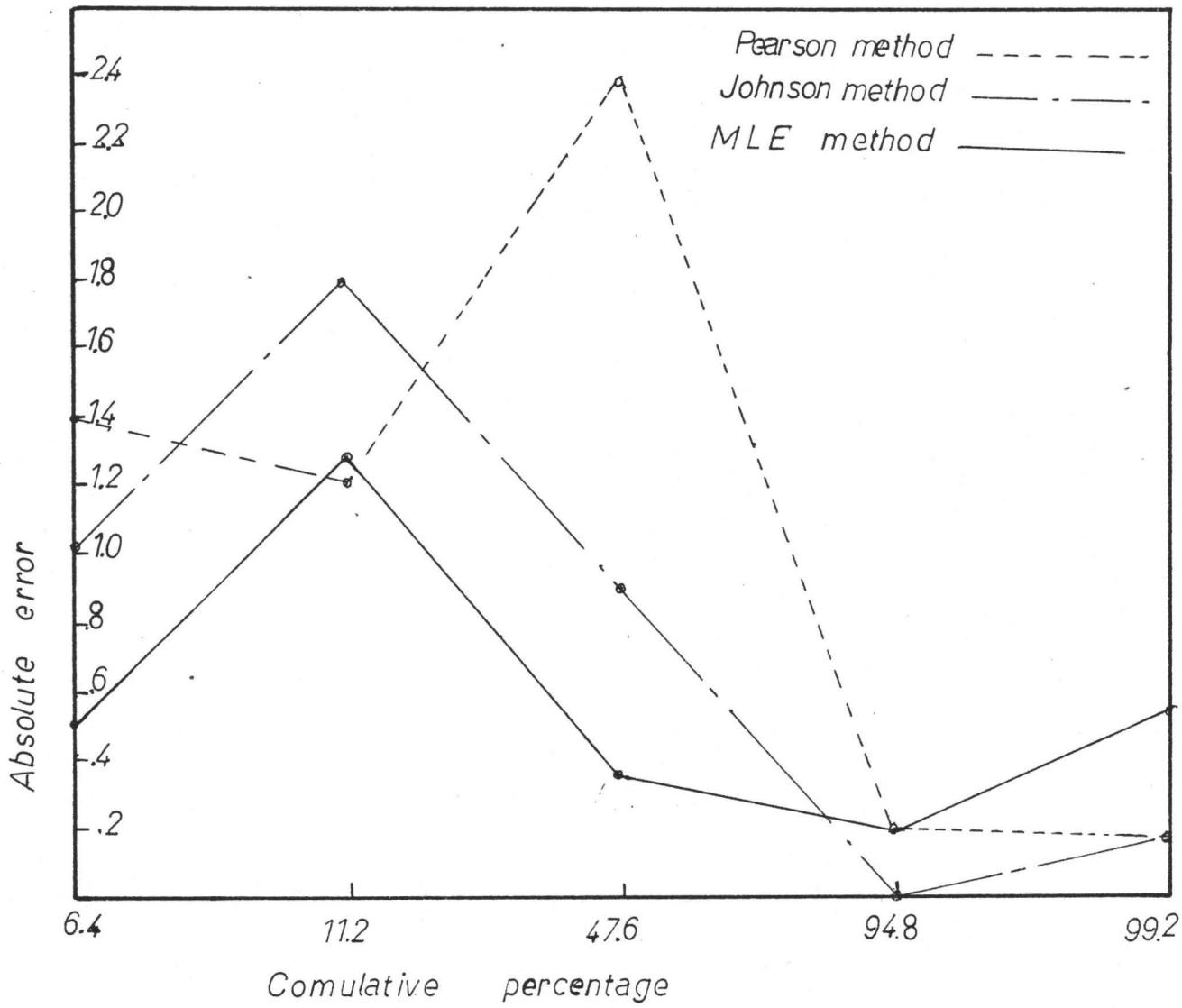


Figure 7.4 Cumulative percentages absolute errors in Pearson, Johnson, and Maximum Entropy Methods.

It is clear from Table 7.4 and Figure 7.4 that the Maximum-Logarithmic Entropy approximation is closer to the actual population than the Johnson approximation or the Pearson approximation.

Table 7.4 Comparison of absolute errors for Johnson, Pearson and Maximum Logarithmic Entropy Methods for calculating the percentages of coefficient of friction data.

Cummulative percentages from actual data	Absolute error in		
	Pearson Method	Johnson Method	Maximum Logarithmic Entropy Method
6.4	1.4	1.0	.52
11.2	1.2	1.8	1.29
47.6	2.4	.9	.36
94.8	.2	0.0	.19
99.2	.2	.2	.55

7.3 THE RESISTORS PROBLEM

This problem is given by Hahn and Shapiro (reference(11), page 215). It is required to fit a distribution to data for the time to complete the manufacture of a part in an automated production process. This time varies from unit to unit because of differences in material quality and hardness. Suppose the minimum cycle time is one half minute for ideal material. The upper bound is two minutes, at which time the material is automatically rejected. The time required for successful completion of 1000 randomly selected units is summarized in the first two columns of Table 7.5.

It was decided to fit a Johnson S_B distribution, first assuming bounds

of 0.5 and 2.0, and then assuming bounds of 0. and 3.0. The result is summarized in the third and fourth columns in Table 7.5. The first four moments were calculated from the first two columns.

$$C_1 = 1.1032000$$

$$C_2 = .04676975$$

$$C_3 = .00599097$$

$$C_4 = .00693529$$

The lower and upper bounds(*) were assumed 0.6 and 1.9 respectively. The Maximum Logarithmic Entropy distribution which satisfies this data is

$$y = \exp(\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3 + \lambda_4 x^4)$$

where

$$\lambda_0 = -50.98$$

$$\lambda_1 = +152.62$$

$$\lambda_2 = -164.91$$

$$\lambda_3 = +78.00$$

$$\lambda_4 = -14.05$$

(*) This is based on the lower limit of the first column, and the higher limit of the last column.

Table 7.5 Comparison between Actual, Predicted Per cent observations by Johnson Method and Maximum Entropy Method Distribution Fits to Production Time for 1000 Randomly Selected Units.

Production Time in Minutes	Actual Per cent of Observations	Predicted Per cent of Observations		
		Johnson Method Assuming bounds of		Maximum Logarithmic Entropy Method
		0.5 and 2.0	0.0 and 3.0	
Less than 0.70	0.9	0.9	1.7	.78
0.70 to 0.79	3.7	4.7	4.3	4.39
0.80 to 0.89	12.6	10.3	9.2	11.72
0.90 to 0.99	18.4	15.2	14.4	18.22
1.00 to 1.09	18.8	17.4	17.7	19.93
1.10 to 1.19	15.8	16.9	17.6	16.20
1.20 to 1.29	12.2	13.9	14.5	11.64
1.30 to 1.39	7.6	10.2	10.1	7.67
1.40 to 1.49	5.0	6.1	5.9	4.78
1.50 to 1.59	2.8	3.1	2.8	2.80
1.60 to 1.69	1.1	1.0	1.2	1.49
1.70 to 1.79	0.9	0.3	0.5	.67
1.80 or more	0.2	0.0	0.1	.23

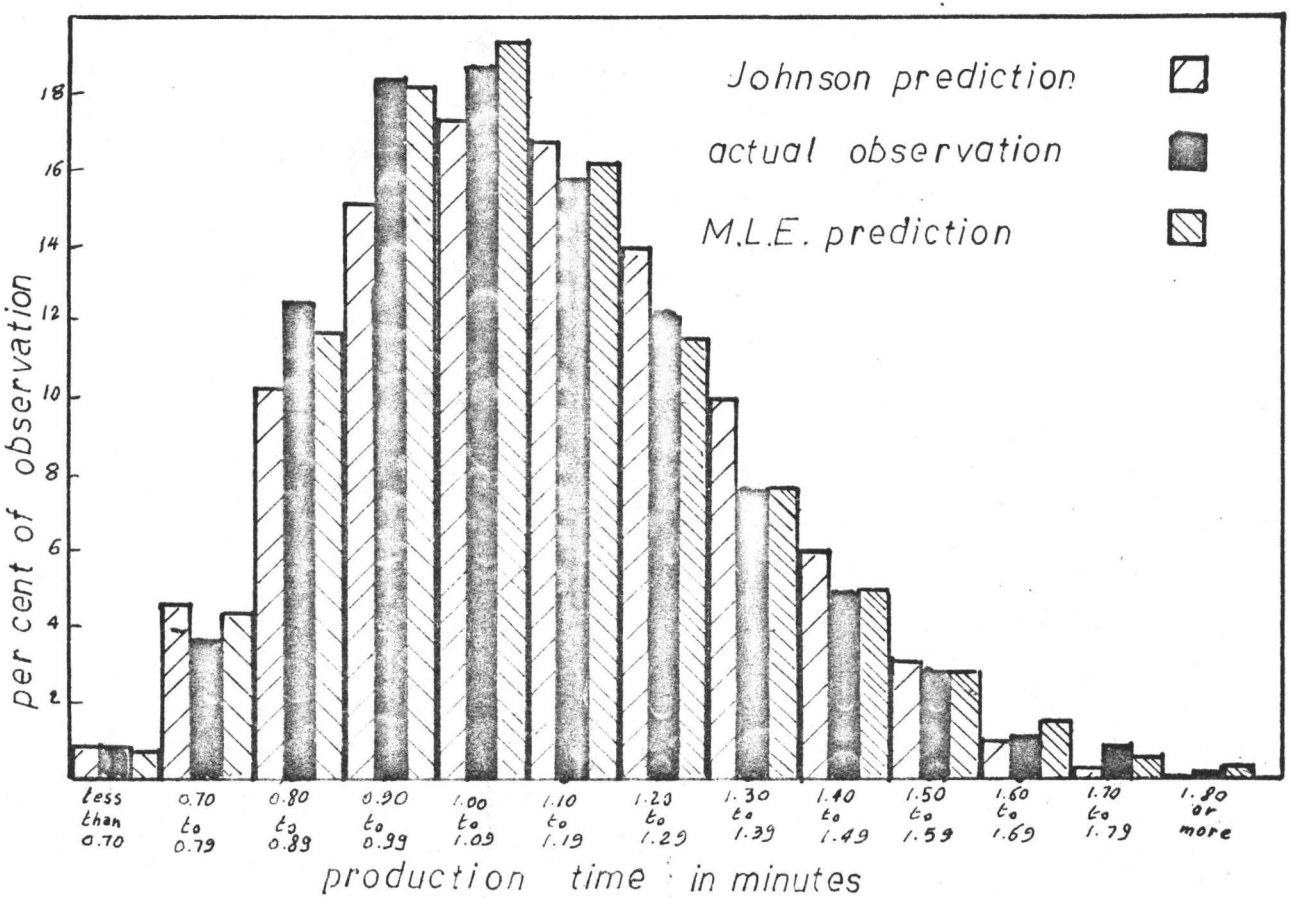


Fig. 7.5 Actual versus predicted Johnson (0.5&2.0 bounds) and predicted M.L.E. per cent of observations of the resistors problem.

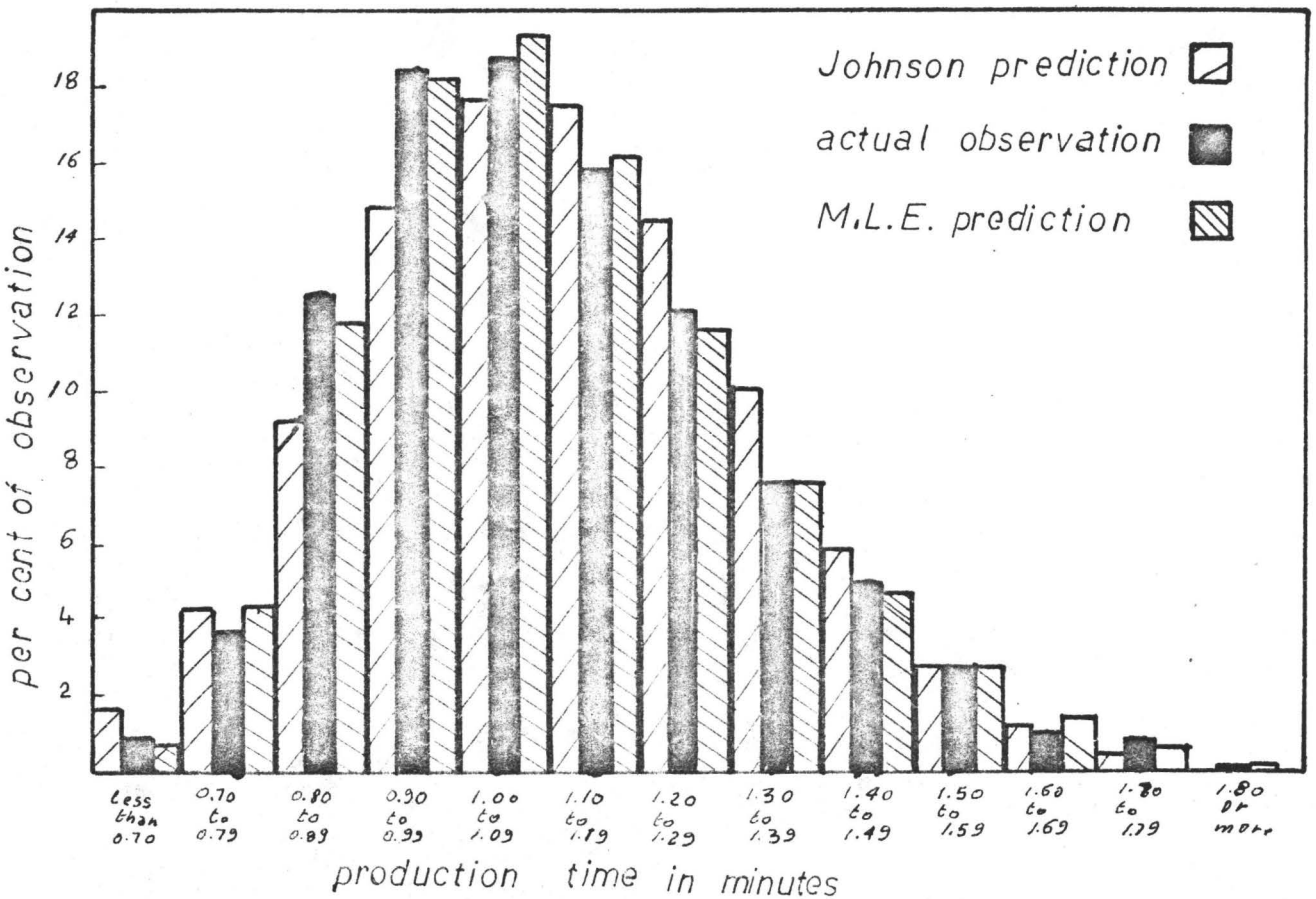


Fig. 7.6 Actual versus predicted Johnson(0.0 & 3.0 bounds) and predicted M.L.E. per cent of observation of the resistors problem

Table 7.6 Comparison of absolute errors in Johnson Method versus absolute errors in Maximum Entropy Method

Actual Per Cent of Observations	Absolute errors in Johnson Method		Absolute errors in Maximum Logarithmic Entropy Method
	Assuming Bounds of 0.5 and 2.0	Assuming Bounds of 0 and 3.0	
0.9	0	.8	.12
3.7	1.	1.	0.69
12.6	2.3	3.4	.88
18.4	3.2	4.0	.18
18.8	1.4	1.1	0.63
15.8	1.1	1.8	.4
12.2	1.7	2.3	.56
7.6	2.6	2.5	.07
5.0	1.1	.9	0.22
2.8	.3	0	0.00
1.1	.1	.1	.39
0.9	.6	.4	.37
0.2	.2	.1	.03

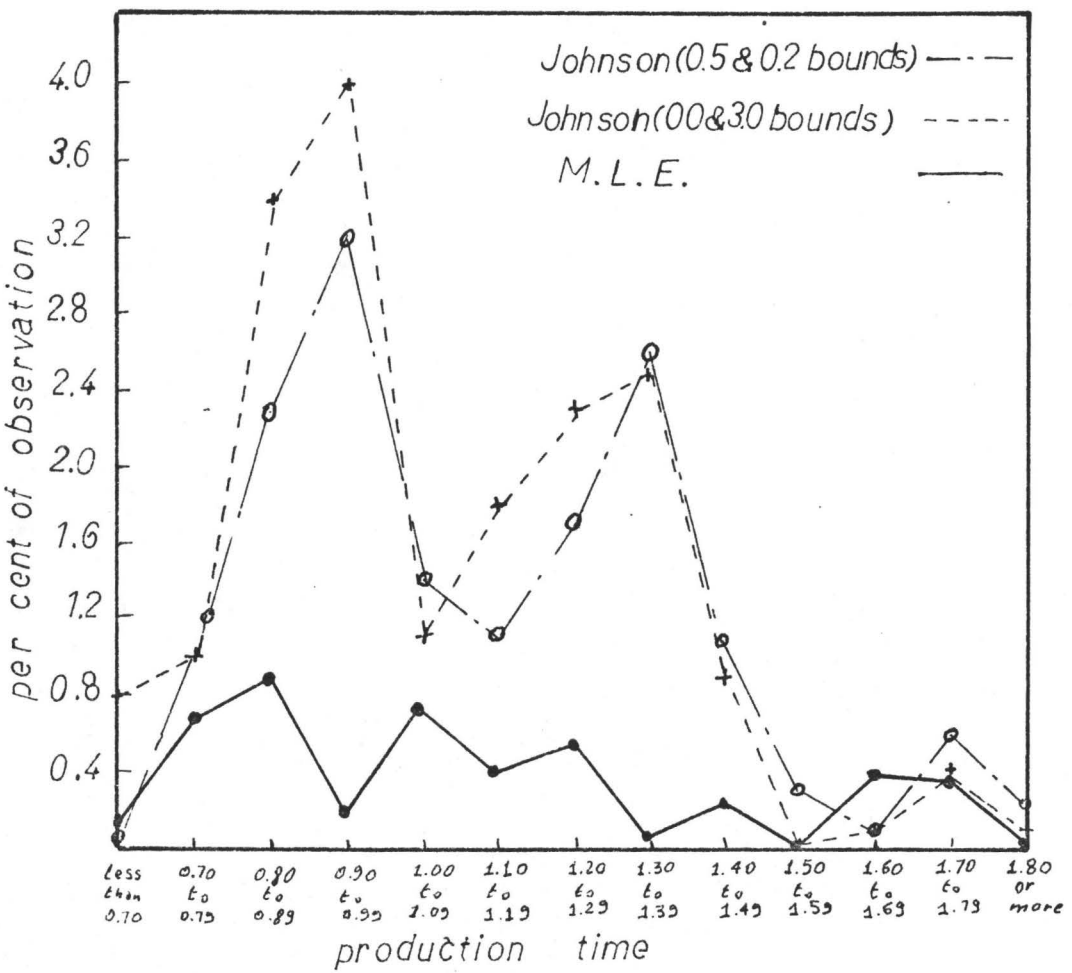


Fig. 7.7 Absolute errors in Johnson method versus absolute errors in M.L.E. method for the resistors problem

The actual and the predicted percent of observations, using the two methods are shown in Table 7.5 and plotted in Figure 7.5 and Figure 7.6. The absolute errors at each point for the two methods are summarized in Table 7.6 and plotted in Figure 7.7.

From Figure 7.7, it is clear that the Maximum Logarithmic Entropy Method is closer to the raw data, and more accurate, than the Johnson Method.

CHAPTER VIII

THE MAXIMUM LOGARITHMIC ENTROPY IN ANALOG PREDICTION UNDER RISK

8.1 GENERAL

We recall from Chapter II that the first moments of y for the general relation

$$y = g(x_1, x_2, \ldots, x_n)$$

can be obtained in terms of the moments of x_i's. We wish to generate the probability distribution of y in terms of its first moments, using the Maximum Logarithmic Entropy Method. The technique is illustrated in the following example.

8.2 THE I-BEAM PROBLEM

The problem as stated by Siddall[1] is that of designing a structure which includes a member in bending. An extruded aluminum I-beam is used having a cross-section shown in Figure 8.1. The density curves are shown in Figures 8.2, 8.3, 8.4, and 8.5 for M the bending moment, y the maximum distance from the neutral axis, I the area moment of inertia and S_y the yield stress of the material. We define strength here as the margin of safety.

$$m = S_y - \frac{MY}{I}$$

The specification value for m is assumed to be zero. We wish to determine the

strength dependability. The density function for M is assumed to be subjective and represented by the Weibull function

$$f(M) = .705 \times 10^{-2} (M - 920{,}000)^{1.33} \exp(-.302 \times 10^{-12} (M - 920{,}000)^{2.33})$$

The density functions for y and I are assumed normal, derived from frequency data.

$$f(y) = \frac{1}{.03 \quad 2\Pi} \exp[-(y - 5.0)^2/2(0.03)^2]$$

$$f(I) = \frac{1}{.76 \quad 2\Pi} \exp[-(I - 163.48)^2/2(7.6)^2]$$

The density function for S_y is assumed to be derived from sampling but has no convenient mathematical model. In the solution sample values must be obtained by table look-up. We wish to determine the density curve for m and the dependability.

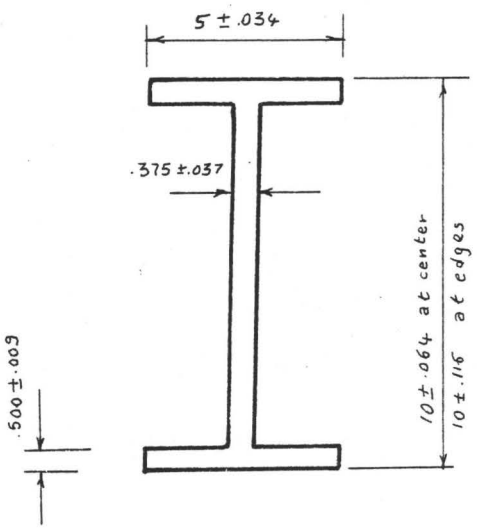


Figure 8.1 Beam Cross-Section

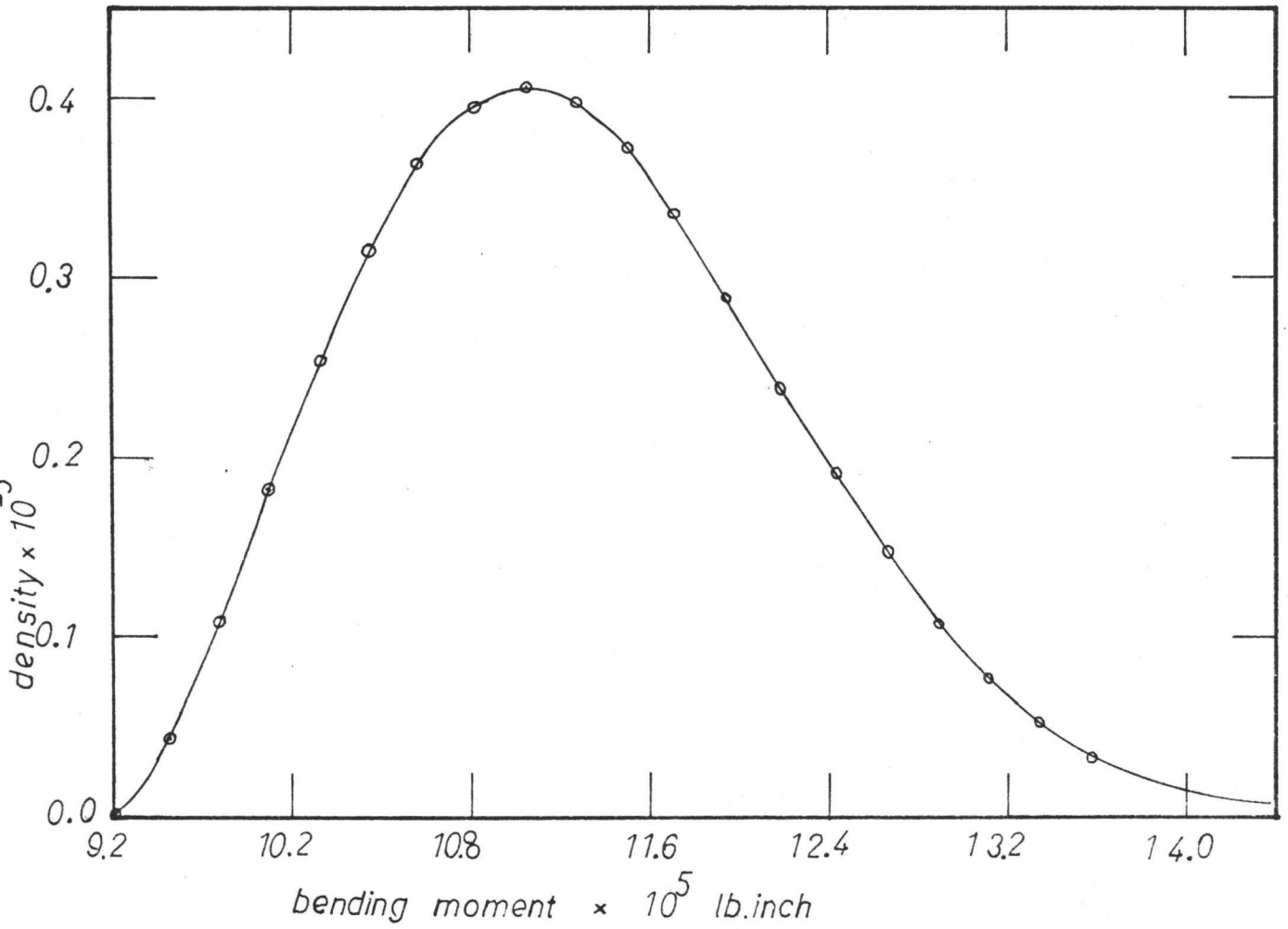


Fig. 8.2 Probability density curve for bending moment M

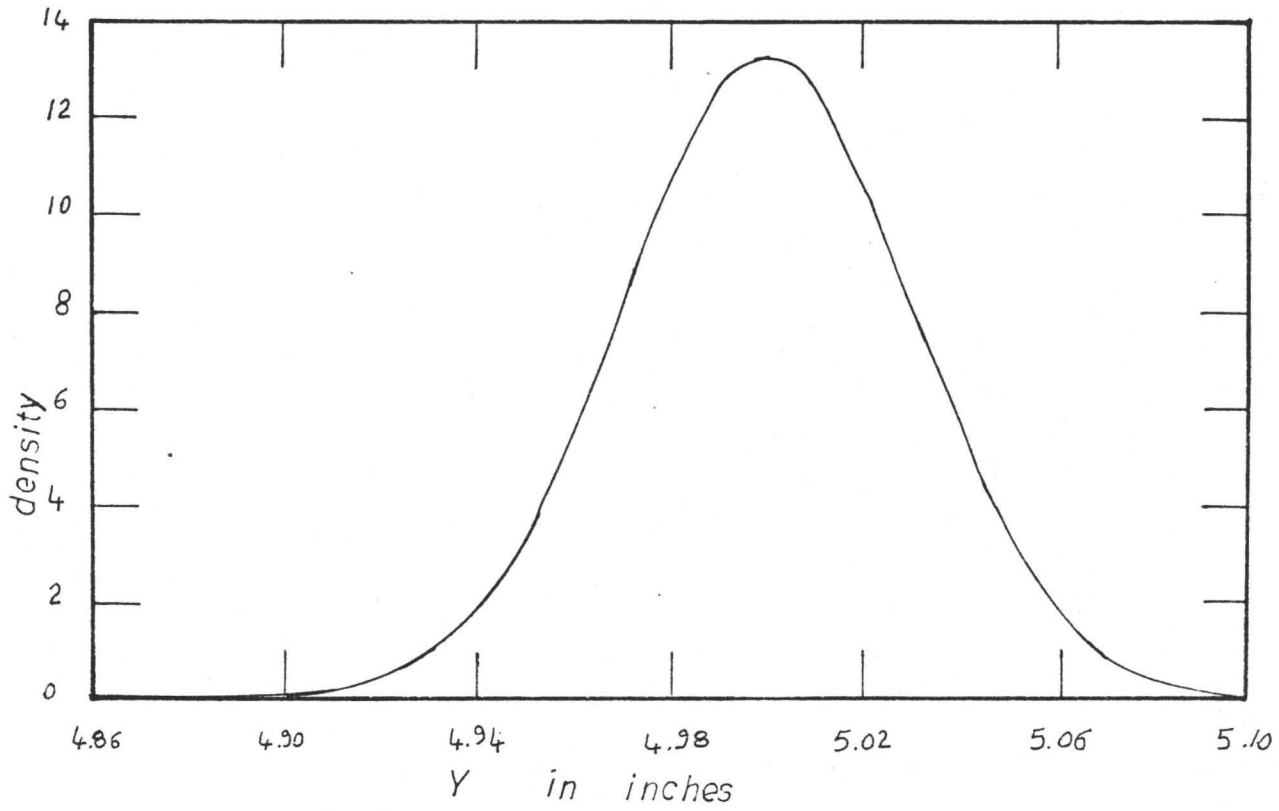


Fig. 8.3 Probability density curve for distance from neutral axis

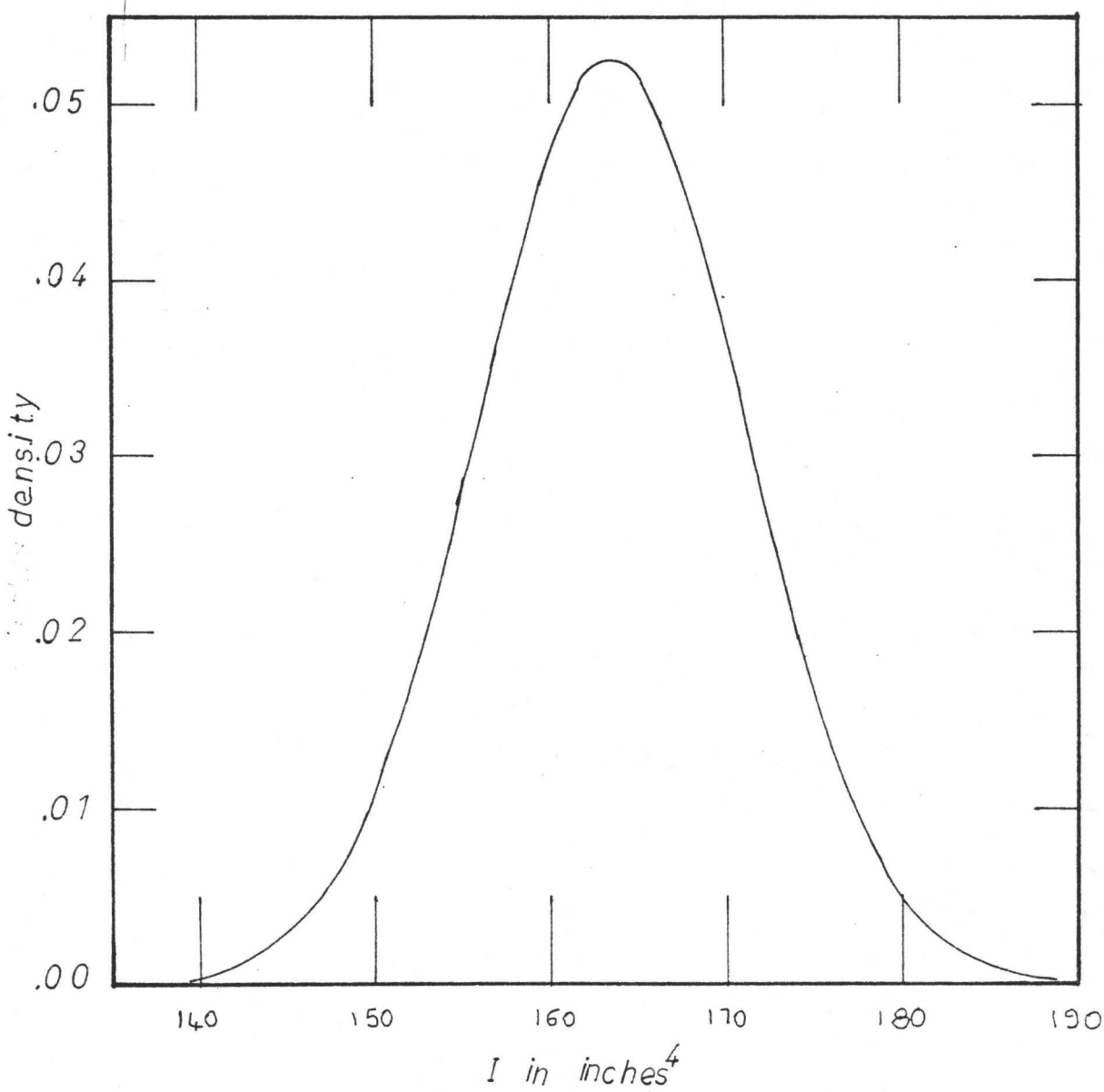


Fig. 84 Probability density curve for moment of inertia of cross section

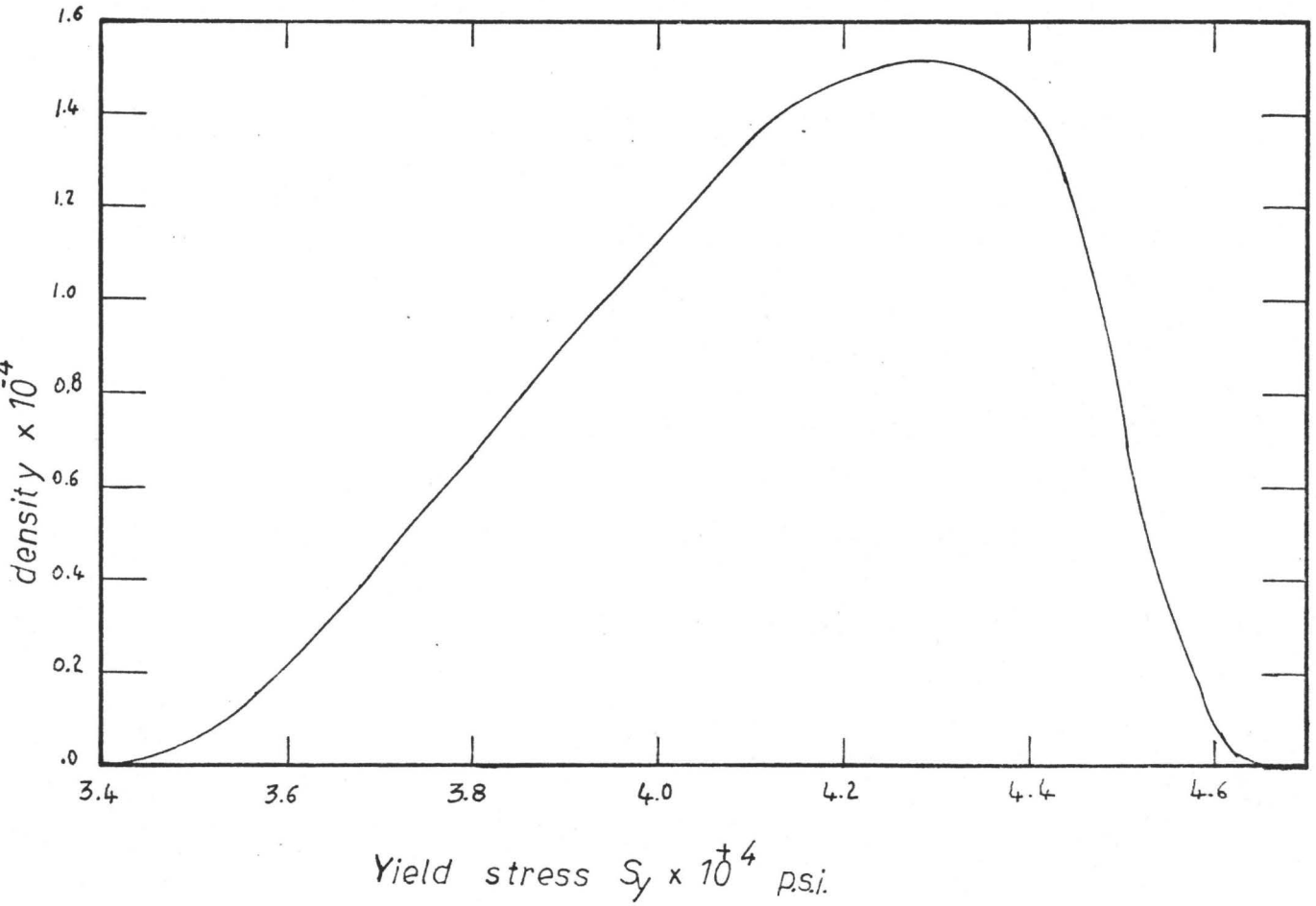


Fig. 8.5 Probability density curve for yield stress of aluminum alloy

The first four moments for S_y, M, Y, and I were calculated using equation (6.1).

$C^1_{S_y} = 4.140651 \times 10^4$ $\quad$ $C^1_M = 1.128693 \times 10^6$

$C^2_{S_y} = 5.785101 \times 10^5$ $\quad$ $C^2_M = 8.897979 \times 10^9$

$C^3_{S_y} = -5.511605 \times 10^9$ $\quad$ $C^3_M = 3.217538 \times 10^{14}$

$C^4_{S_y} = 8.001181 \times 10^{13}$ $\quad$ $C^4_M = 2.180712 \times 10^{20}$

$C^1_y = 5.0$ $\quad$ $C^1_I = 1.634987 \times 10^2$

$C^2_y = 8.907379 \times 10^{-4}$ $\quad$ $C^2_I = 5.697391 \times 10$

$C^3_y = +0.0$ $\quad$ $C^3_I = 8.271875$

$C^4_y = 2.312377 \times 10^{-6}$ $\quad$ $C^4_I = 9.410815 \times 10^3$

The upper and lower bounds were assumed from the figures as follows

$XMIN_{S_y} = 3.4 \times 10^4$ $\quad$ $XMAX_{S_y} = 4.7 \times 10^4$

$XMIN_M = 9.2 \times 10^5$ $\quad$ $XMAX_M = 1.44 \times 10^6$

$XMIN_y = 4.9$ $\qquad$ $XMAX_y = 5.1$

$XMIN_I = 1.4 \times 10^2$ $\qquad$ $XMAX_I = 1.9 \times 10^2$

Using a truncated Taylor's series expansion as in Section (2.53), the first four moments of m are

$$C_m^1 = 6.816094 \times 10^3$$

$$C_m^2 = 1.668383 \times 10^7$$

$$C_m^3 = -1.463594 \times 10^{10}$$

$$C_m^4 = 7.974988 \times 10^{14}$$

The upper and lower bounds for m are

$$XMIN_m = -1.845714 \times 10^4$$

$$XMAX_m = 2.327369 \times 10^4$$

The Maximum Logarithmic Entropy Distribution (the density curve) which satisfy the above conditions is:

$$f(m) = \exp(\lambda_0 + \lambda_1 m + \lambda_2 m^2 + \lambda_3 m^3 + \lambda_4 m^4)$$

where

$$\lambda_0 = -1.0587 \times 10$$

$$\lambda_1 = 3.5265 \times 10^{-4}$$

$$\lambda_2 = -2.4887 \times 10^{-8}$$

$$\lambda_3 = +5.7094 \times 10^{-13}$$

$$\lambda_4 = -4.7247 \times 10^{-17}$$

The dependability, which is simply equal to the probability that $m > 0$ is equal to .9449. The computer time for this problem (CP time) = 14 sec. Cost in $ (based on $600/hr) = 2.33. The results shown above are obtained by using the computer package 'DECII' given in Appendix B.

For the same above first four moments the dependability by using Pearson's method is 95.88%. By using the transformation of variables technique the dependability is 94.91%. We consider this result is the exact one as theoretically this technique does not depend on approximation, however, this method is applicable for independent variables only (which is our case).

By using the Monte-Carlo Simulation technique the result varies with the sampe size (N), the following solutions have been provided by Professor J. N. Siddall.

N	Dependability	Cost in $
1000	94.8	8.42
5000	95.3	31.47
10000	94.9	60.71
12000	95.1	72.32
14000	95.0	84.10
16000	95.0	95.73

The error in the case of the Maximum Logarithmic Entropy is 0.42. However, in Monte-Carlo the error varies from -.39 to .11 and is dependent on the sample size, which is nearly equal to the error in the new technique. But the new method is considerably better from the point of view of cost ($2.33 vs $95.73). Also the accuracy in the new technique could be improved by considering the fifth or the sixth moment. In Pearson's method the error goes up to .97 which is double the error in the new technique.

CHAPTER IX

GENERAL DISCUSSION

The algorithm of maximization of the logarithmic entropy function subject to the first n moments as constraints, as introduced in this thesis, can be regarded as an important contribution to analytical decision theory. It is the first step towards the meaningful application of the Maximum Entropy Principle to decision theory. In addition this thesis contains the first valid experimental proof of the principle mentioned above. It lacks any theoretical proof , as do many principles and theories (for example, Newton's law), but most of these principles and theories have empirical proof, or, in other words, there are data or results which support these theories. However, when Jaynes introduced the principle, which is an application of the Shannon's entropy function, he did not support it by a theoretical proof or by applications. Later, Tribus illustrated that the principle worked for special cases (the uniform, the gamma, and the normal distributions). The thesis illustrates the validity of the principle in a much more general sense. Although the application of the principle in this thesis was limited to a special type of constraint function, the moment function, this limitation does not affect the validity of the principle. Other types of functions can be easily handled, which will be considered later.

Besides illustrating the generality of Jayne's Principle, this work has achieved the following advantages in comparison with other existing empirical methods:

1) In comparison with the Monte Carlo method, it is clear from Chapter VIII that the new algorithm is more accurate with less computation time.

2) In comparison with Pearson's method the new algorithm is more accurate. This can be observed from Chapter VII. No computation time comparison has been made since the Pearson's method is not available in computer language. Pearson's method limits itself to the first four moments which mean that if only the first three moments are available, Pearson's method cannot predict the distribution. The same applies for the first five moments, if a more accurate result is required. The new algorithm does not have the disadvantage.

As was demonstrated in Chapter VI, an error occurs and increases at higher moments, in addition to the error from calculating the higher moments. To eliminate this increasing error in the final solution, other types of functions can be suggested to replace the moment function. One such function has the following form

$$C_1 = < \log (x) >$$

$$C_i = < \log (x) - C_1)^i > \quad i = 2, 3, \ldots, n$$

However, many different kinds of functions could be suggested, the best being the one which gives the least error in the final solution. This would appear to be a fruitful source of future work in this area, using the same approach as was used in this thesis.

REFERENCES

1. Siddall, J. N., Analytical Decision Making in Engineering, Prentice-Hall, 1972.
2. Tribus, M., Rational Descriptions, Decisions and Designs, Pergamon Press, 1969.
3. Kendall, M. G. and A. Stuart, The Advanced Theory of Statistics, Vol. 1, Hafner Publishing Company, New York, 1958.
4. Hahn, G. J. and S. S. Shapiro, Statistical Models in Engineering, John Wiley & Sons, Inc., New York, 1967.
5. Johnson, N. L., "Systems of Frequency Curves Generated by Methods of Translation", Biometrika, 36, 149 (1949).
6. Johnson, N. L., E. Nixon and D. E. Amos, "Tables of Percentage Points of Pearson Curves, for given β_1, β_2, Expressed in Standard Measure", Biometrika, 50, 459 (1963).
7. Elderton, W. P., Frequency Curves and Correlation, 4th ed., Cambridge University Press, Cambridge, 1953.
8. Chacko, G. K., Applied Statistics in Decision-Making, American Elsevier Publishing Company, Inc., 1971.
9. Mize, J. H., and J. G. Cox, Essentials of Simulation, Prentice-Hall, Inc., Englewood Cliffs, N. J. (1968).
10. Shannon G. E., A Mathematical Theory of Communication, The Bell System Technical Journal, Vol XXVII, July, 1948, No. 3.
11. Tribus, M., Thermostatics and Thermodynamics, D. Van Nostrand Co., 1961.

12. Mathai, A. M., and P. N. Rathie, Axiomatic foundations of some basic concepts in information theory and statistics, McGill University, September, 1971.

13. Jaynes, E. T., "Prior Probabilities", IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, September, 1968.

14. Behara, M., and P. Nath, "Additive and non-additive entropies of finite measurable partitions", McMaster Mathematical Report, No. 31, September, 1970.

15 Tribus M., "The use of the Maximum Entropy Estimate in the Estimation of Reliability", in Recent Developments in Information and Decision Processes Edited by R. E. Machol, The Macmillan Company, 1962.

16. Reza, F. M., An introduction to information theory, McGraw-Hill Book Co., Inc., 1961.

17. Haugen, E., Probabilistic Approaches to Design, John Wiley & Sons, Inc., 1968.

18. Korn and Korn, Mathematical Handbook for Scientists and Engineering, McGraw-Hill Book Co., Inc., 1961.

19. Spiegel, M. R., Mathematical Handbook of formulas and tables, McGraw-Hill Book Co., Inc., 1968.

20. Jacobson D. H., and W. Oksman, An Algorithm that minimizes homogeneous functions of N variables in N + 2 iterations and rapidly minimized general functions, Technical Report No. 618, Harvard University, Cambridge, Mass., October, 1970.

21. Fletcher, R., A new approach to variable metric algorithms, the Computer Journal, Volume 13, Number 3, August, 1970.

APPENDIX A

Systems of Analytical Probability Distributions Approximated by the Maximum Logarithmic Entropy Method

Table A.1. List of Analytical Distributions with corresponding Figures and Tables numbers.

Symbol	Distributions name and parameter values	Table Number	Figure Number
G1	Gamma, $\eta = 3.0$, $\lambda = 1.$	A.2	A.1
G2	Gamma, $\eta = 3.0$, $\lambda = 5.0$	A.3	A.2
G3	Gamma, $\eta = 1.0$, $\lambda = 1.0$	A.4	A.3
B1	Beta, $\eta = 5.0$, $\gamma = 1.5$	A.5	A.4
B2	Beta, $\eta = 2.0$, $\gamma = .8$	A.6	A.5
B3	Beta, $\eta = 2.0$, $\gamma = 1.0$	A.7	A.6
B4	Beta, $\eta = .5$, $\gamma = .5$	A.8	A.7
B5	Beta, $\eta = 5.0$, $\gamma = 5.0$	A.9	A.8
B6	Beta, $\eta = 3.0$, $\gamma = 1.5$	A.10	A.9
B7	Beta, $\eta = 3.0$, $\gamma = 3.0$	A.11	A.10
B8	Beta, $\eta = 2.0$, $\gamma = 2.0$	A.12	A.11
B9	Beta, $\eta = 1.5$, $\gamma = 3.0$	A.13	A.12
B10	Beta, $\eta = 1.5$, $\gamma = 5.0$	A.14	A.13
B11	Beta, $\eta = 2.0$, $\gamma = .5$	A.15	A.14
W1	Weibull, $\eta = 1.0$, $\sigma = 1.0$	A.16	A.15
W2	Weibull, $\eta = 4.0$, $\sigma = 1.0$	A.17	A.16
W3	Weibull, $\eta = 2.0$, $\sigma = 1.0$	6.1	6.1 to 6.5
W4	Weibull, $\eta = .5$, $\sigma = 1.0$	A.18	A.17
E	Exponential, $\lambda = 1.0$	A.19	A.18
U	Uniform, $\mu_0 = 1.0$, $\mu_1 = 1.0$	A.20	A.19 & A.20
N	Normal, $\mu = 4.0$, $\sigma = 1.0$	A.21	A.21

Symbol	Distributions name and parameter values	Table Number	Figure Number
TN	Truncated Normal, $\mu = 0, \sigma = 2.0$	A.22	A.22 & A.23
HN	Half Normal, $\sigma = 5.0$	A.23	A.24
R1	Rayleigh, $\sigma^2 = 2.0$	A.24	A.25
R2	Rayleigh, $\sigma^2 = 1,0$	A.25	A.26
R3	Rayleigh, $\sigma^2 = .5$	A.26	A.27
C1	Cauchy, $\mu = 0, \sigma = 1.0$	A.27	A.28
C2	Cauchy, $\mu = 0, \sigma = 1.0$	A.28	A.29
LN1	Log-Normal, $\mu = 1.0, \sigma^2 = 1.0$	A.29	A.30
LN2	Log-Normal, $\mu = .3, \sigma^2 = 1.0$	A.30	A.31
LN3	Log-Normal, $\mu = 0.0, \sigma^2 = 1.0$	A.31	A.32
LN4	Log-Normal, $\mu = 0.0, \sigma^2 = ,3$	A.32	A.33
LN5	Log-Normal, $\mu = 0.0, \sigma^2 = .1$	A.33	A.34

GAMMA DISTRIBUTION (G1)

$$f(x) = \begin{cases} \frac{\lambda^{\eta}}{\eta} x^{\eta - 1} e^{-\lambda x} & \eta = 3.0,\ \lambda = 1.0,\ x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values = 0.59989 0.11974 0.04731

0.07006 0.07533 0.10915

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 0.74748 3.12731

Upper and lower bounds: 8.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.2 Common area between Gamma Distribution (G1) and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
2	90.27
3	94.68
4	96.88
5	97.99

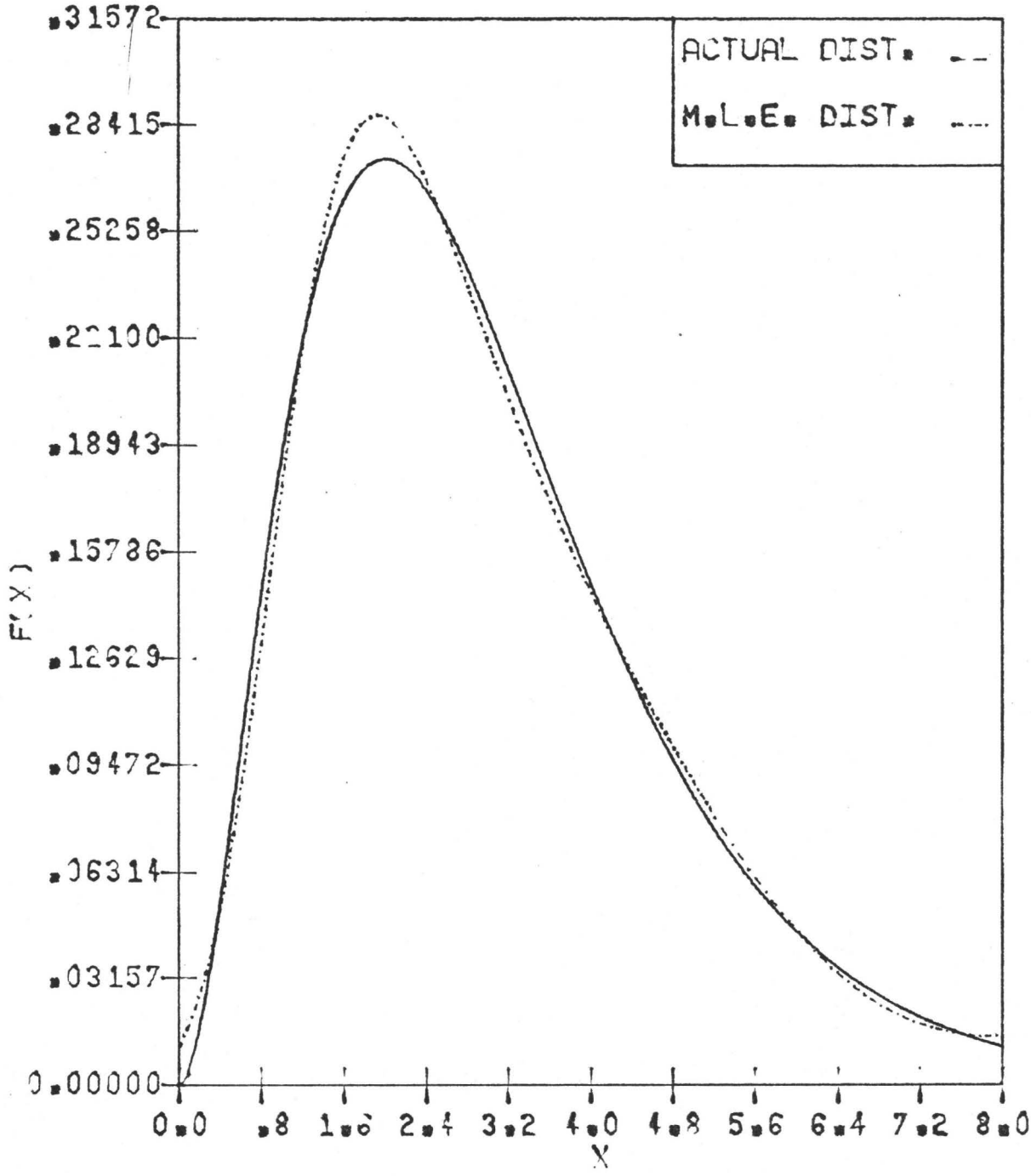


Figure A.1 Approximating Gamma Distribution (G1), (η = 3.0, λ = 1.0) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central moments values: 2.91292 2.46997 2.90160 19.07892 53.47114

λ values (for M.L.E. Distribution): - 4.4888 + 4.68476 - 2.45213 + .58318 - .067767 + .003013

Percentage area in common between the two distributions = 97.99

GAMMA DISTRIBUTION (G2)

$$f(x) = \begin{cases} \frac{\lambda^{\eta}}{\eta} x^{\eta - 1} e^{-\lambda x} & \eta = 3.0,\ \lambda = 5.0,\ x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: .59989 .119736 .04731

.07006 .07533

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 1.14184 4.88698

Upper and lower bounds: 3.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.3 Common area between Gamma Distribution (G2), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
2	88.95
3	92.41
4	95.62
5	96.81

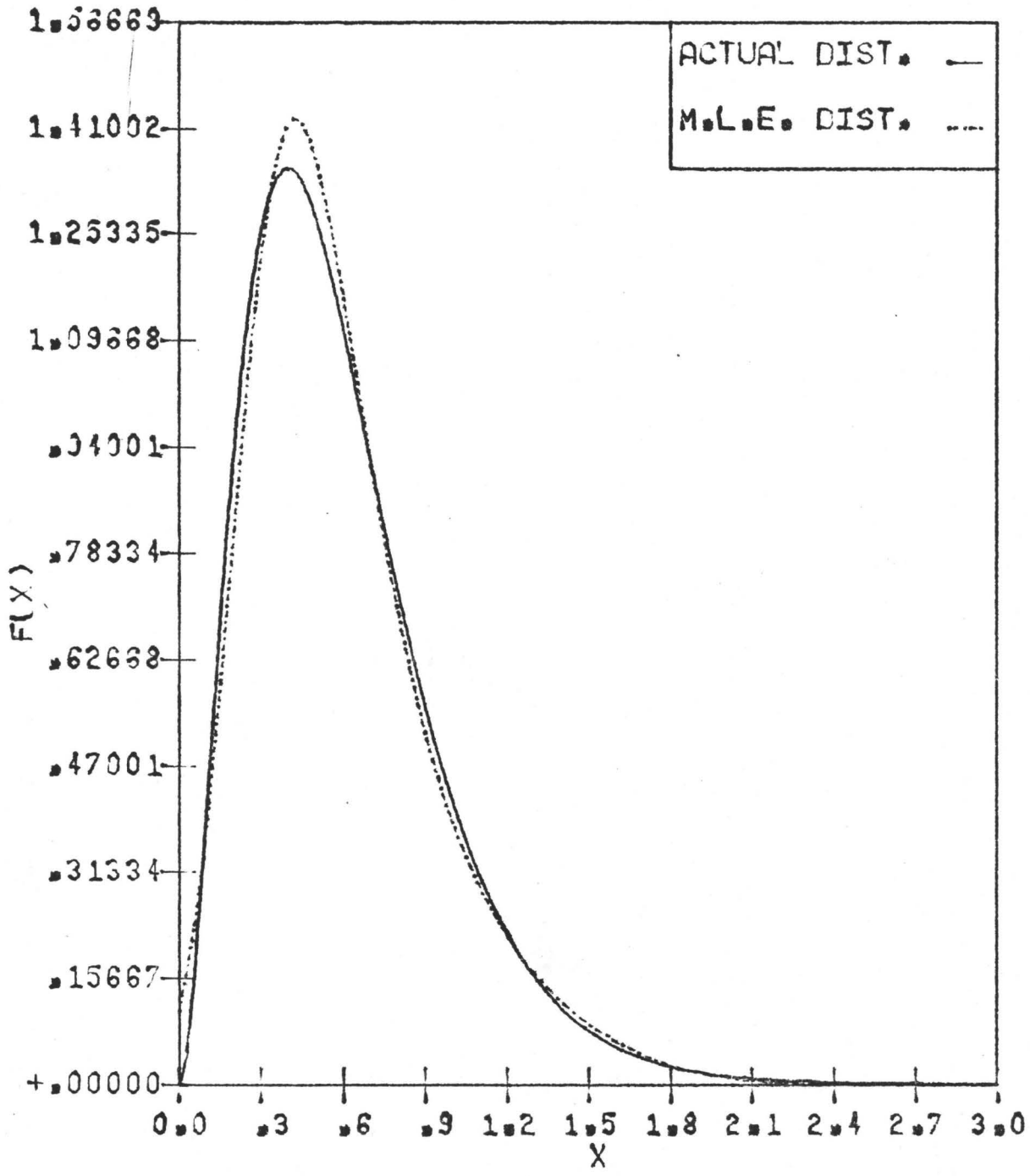


Figure A.2 Approximating Gamma Distribution (G2), (η = 3.0, λ = 5.0) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .59989 .119736 .04731 .07006 .07533

λ values (for M.L.E. Distribution): - 2.26178 + 15.32115 30.21764 + 23.94046 - 8.96673 + 1.24403

Percentage area in common between the two distributions = 96.81

GAMMA DISTRIBUTION (G3)

$$f(x) = \begin{cases} \frac{\lambda^{\eta}}{\eta} x^{\eta - 1} e^{-\lambda x} & \eta = \lambda = 1.0,\ x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: .96608 .82925 1.13471

3.61955 10.00447 31.615521

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 1.5026 5.2635

Upper and lower bounds: 5.00 1.00

Tupe of curve: J shape

Table A.4 Common area between Gamma Distribution (G3), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments

Number of known first moments	Percentage of area in common between the two curves
2	100.00
3	99.99
4	99.97
5	99.94
6	99.90

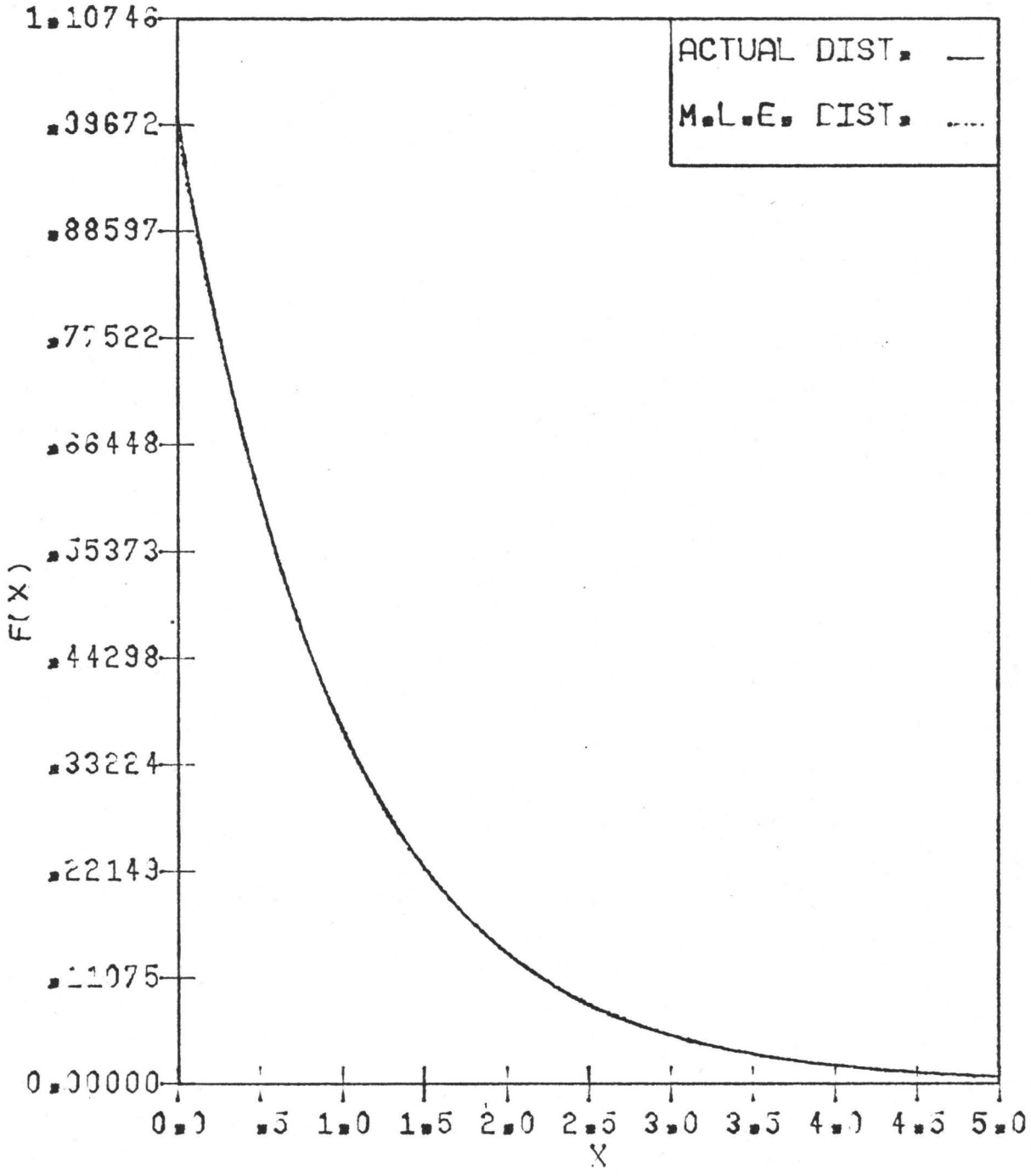


Figure A.3 Approximating Gamma Distribution (G3), ($\eta = \lambda = 1.0$) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .96608 .82925 1.13471 3.61955 10.00447

λ values (for M.L.E. Distribution): .003646 - .976842 - .038436 + .023483 - .005896 + .000516

Percentage area in common between the two distributions = 99.94

BETA DISTRIBUTION (B1)

$$f(x) = \begin{cases} \dfrac{\Gamma(\eta + \gamma)}{\Gamma(\eta)\,\Gamma(\gamma)}\, x^{\gamma - 1}(1 - x)^{\eta - 1}, & \eta = 5.0,\ \gamma = 1.5,\ 0 \leq x \leq 1 \\ 0, \text{ elsewhere} & \end{cases}$$

Central moment values: .23103 .02364 .00300

.00183 .00058

Standardized moment measures ($\sqrt{\beta_1}$, β_2): .82502 3.2803

Upper and lower bounds: 1.00 0.00

Tupe of curve: $\cap$ shape (Bell Shaped)

Table A.5 Common area between Beta Distribution (B1), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	96.09
4	97.55
5	97.89

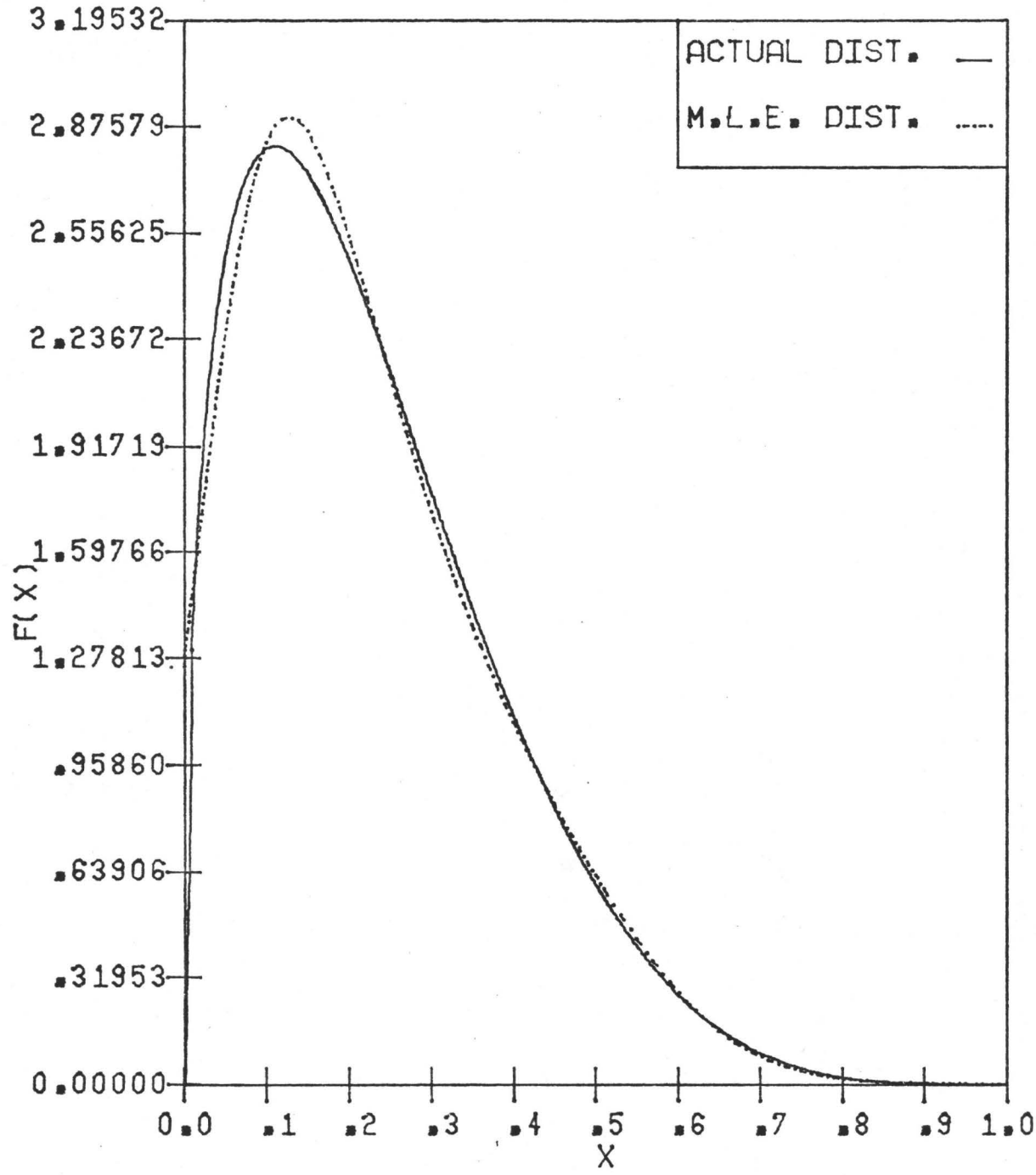


Figure A.4 Approximating Beta Distribution (B1), (η = 5.0, γ = 1.5) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .231027 .023635 .002998 .0018324 .0005786

λ values (for M.L.E. Distribution): + .2255 + 15.9965 - 100.1084 + 240.0444 - 274.1477 110.8254

Percentage area in common between the two distributions = 99.89

BETA DISTRIBUTION (B2)

$$f(x) = \begin{cases} \dfrac{\Gamma(\eta + \gamma)}{\Gamma(\eta)\,\Gamma(\gamma)}\, x^{\gamma - 1}(1 - x)^{\eta - 1}, & \eta = 2.0,\ \gamma = 0.8,\ 0 \leqslant x \leqslant 1 \\ 0, \text{ elsewhere} & \end{cases}$$

Central moment values: .287151 .053566 .009524

.007746 .003086 .0018559

Standardized moment measures ($\sqrt{\beta_1}$, β_2): .7682 2.6996

Upper and lower bounds: 1.00 0.00

Type of curve: J shape

Table A.6 Common area between Beta Distribution (B2), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	97.54
4	97.55
5	98.46
6	98.48

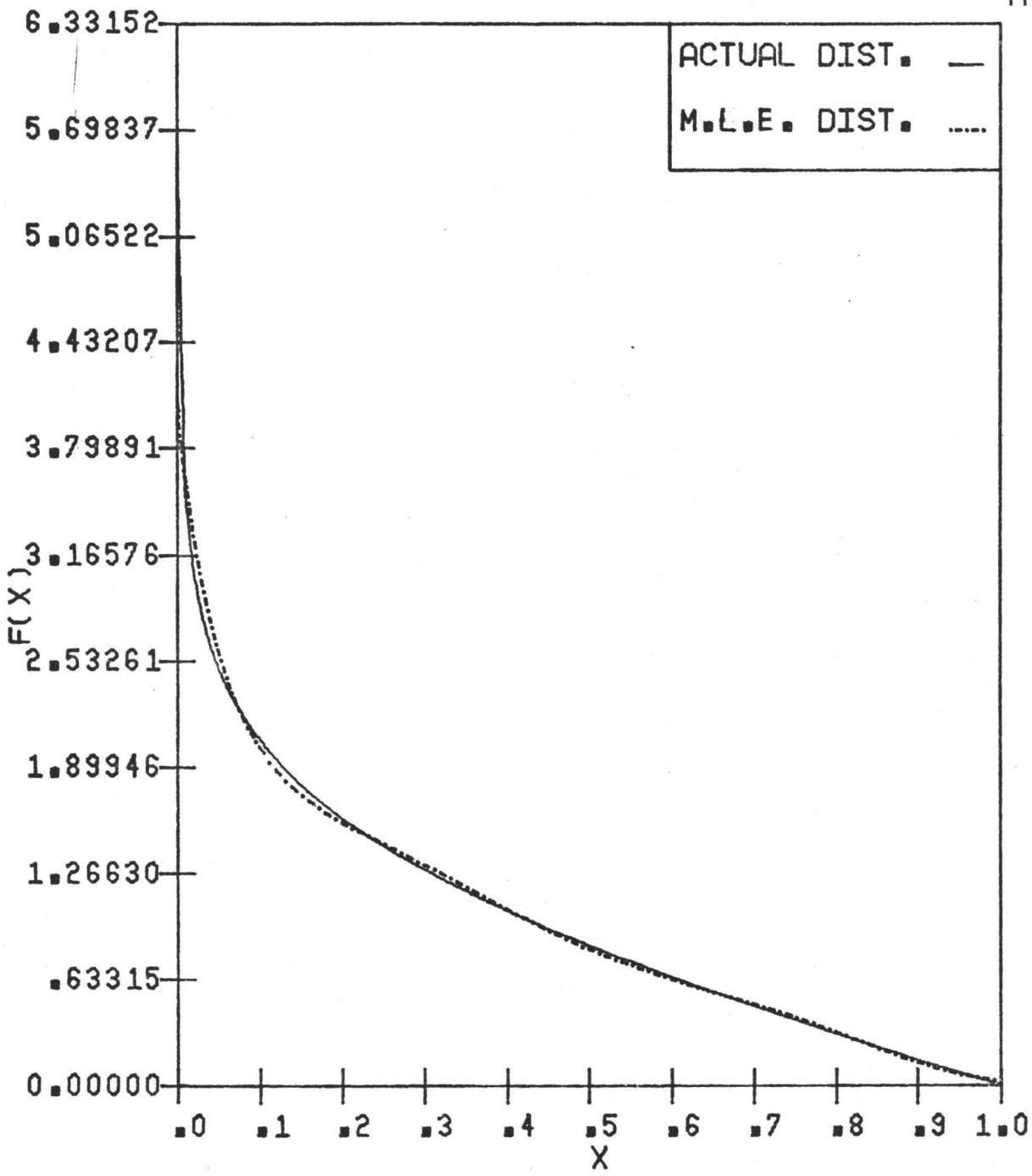


Figure A.5 Approximating Beta Distribution (B2), (η = 2.0, γ = .8), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first six moments.

Central Moment values: .28715 .05357 .00952 .00775 .00309 .00186

λ values (for M.L.E. Distribution): + 1.3972 - 11.4708 + 60.4860 - 178.7122 + 258.3163 - 173.3041 + 39.6237

Percentage area in common between the two distributions = 98.48

BETA DISTRIBUTION (B3)

$$f(x) = \begin{cases} \dfrac{\Gamma(\eta + \gamma)}{\Gamma(\eta)\,\Gamma(\gamma)}\, x^{\gamma - 1}(1 - x)^{\eta - 1}, & \eta = 2.0,\ \gamma = 1.0,\ 0 \leq x \leq 1 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: .33400 .05545 .00739

-.00738 .00234 .00151

Standardized moment measures ($\sqrt{\beta_1}$, β_2): .5657 2.4000

Upper and lower bounds: 1.00 0.00

Type of curve: triangular

Table A.7 Common area between Beta Distribution (B3), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	98.44
4	99.01
5	99.38
6	99.55

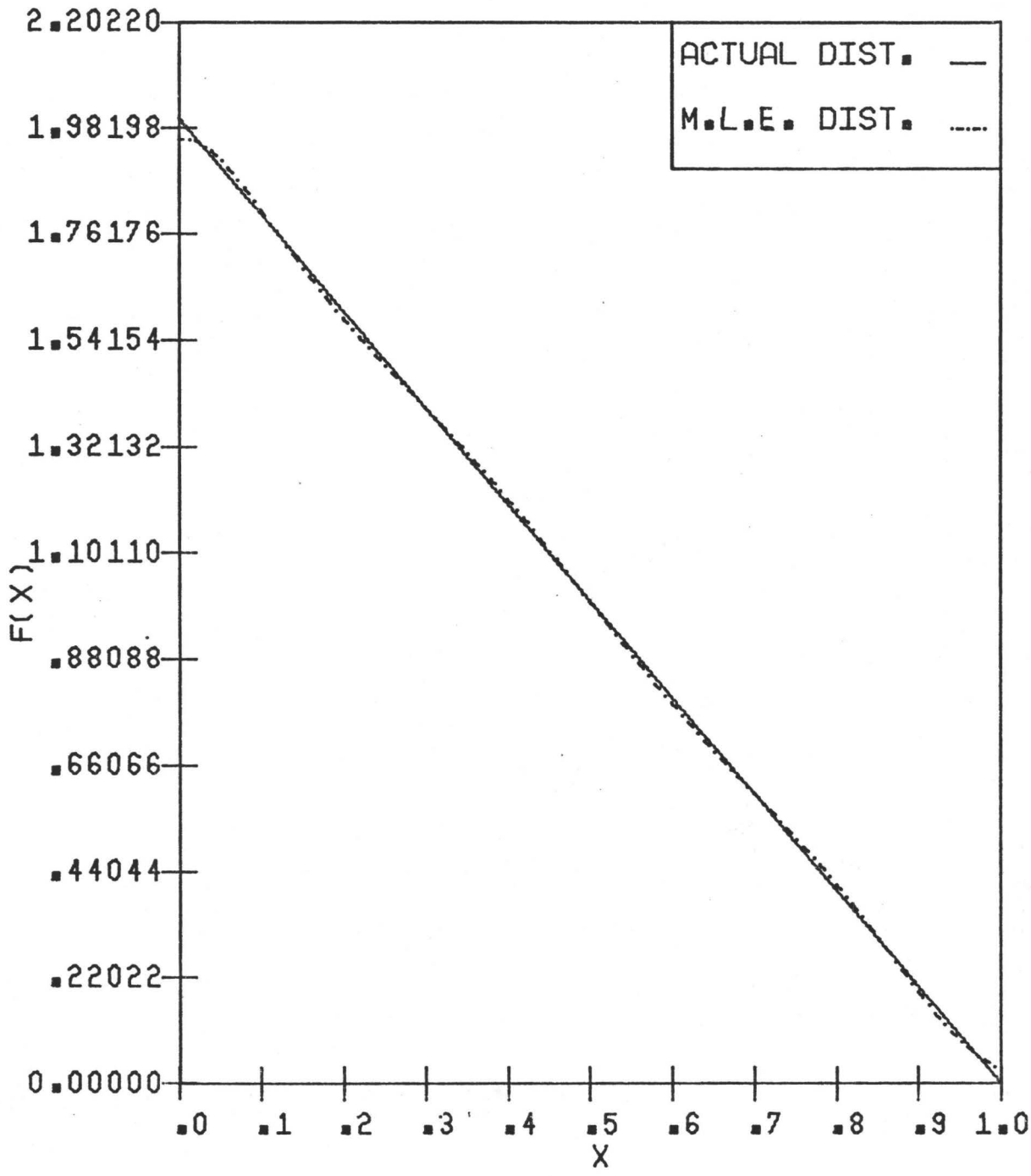


Figure A.6 Approximating Beta Distribution (B3), (η = 2.0, γ = 1.0) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first six moments.

Central Moments values: .33400 .05545 .00739 -.00738 .00234 .00151

λ values (for M.L.E. Distribution): + .67267 + .22341 - 16.18761 78.87925 - 185.39258 + 201.18329 - 82.96117

Percentage area in common between the two distributions = 99.55

BETA DISTRIBUTION (B4)

$$f(x) = \begin{cases} \dfrac{\Gamma(\eta + \gamma)}{\Gamma(\eta)\,\Gamma(\gamma)}\, x^{\gamma - 1}(1 - x)^{\eta - 1}, & \eta = \gamma = .5,\ 0 \leqslant x \leqslant 1 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: .50000 .12223 0.00000

.02258 0.00000

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 0.00 1.51

Upper and lower bounds: 1.00 0.00

Type of curve: *U* shape

Table A.8 Common area between Beta Distribution (B4), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	88.78
4	92.15
5	92.15

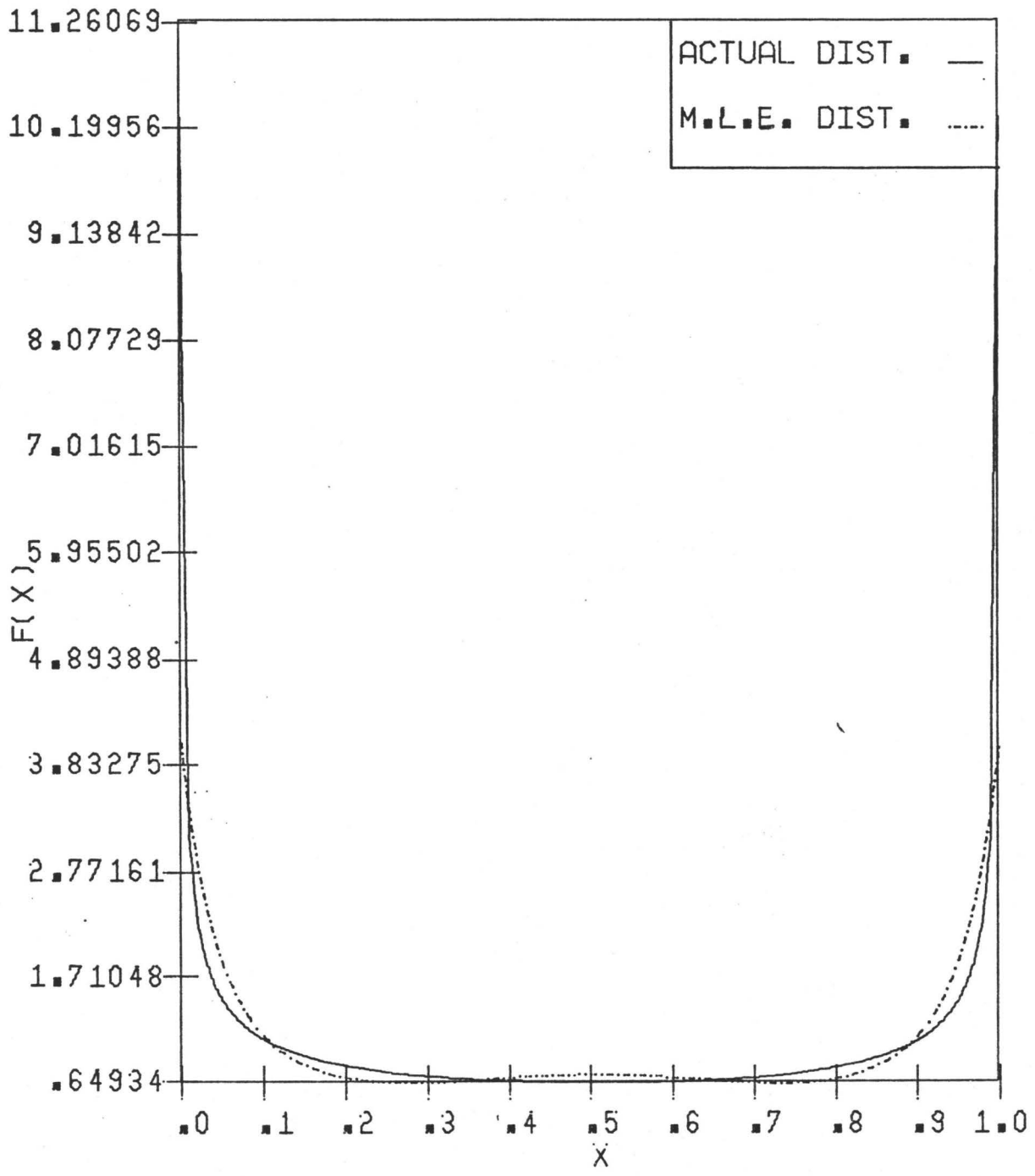


Figure A.7 Approximating Beta Distribution (B4), (η = .5, γ = .5) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .50000 .12223 0.00000 .02258 .00000

λ values (for M.L.E. Distribution): + 1.4178 - 18.6369 + 65.5595 - 93.74824 46.70689 119.35532

Percentage area in common between the two distributions = 92.15.

BETA DISTRIBUTION (B5)

$$f(x) = \begin{cases} \dfrac{\Gamma(\eta + \gamma)}{\Gamma(\eta)\,\Gamma(\gamma)}\, x^{\gamma - 1}(1 - x)^{\eta - 1}, & \eta = \gamma = 5,\ 0 \leq x \leq 1 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: .5000 .02273 0.0000

.001311 .0000

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 0.00 2.54

Upper and lower bounds: 1.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.9 Common area between Beta Distribution (B5) and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	96.65
4	99.26
5	99.26

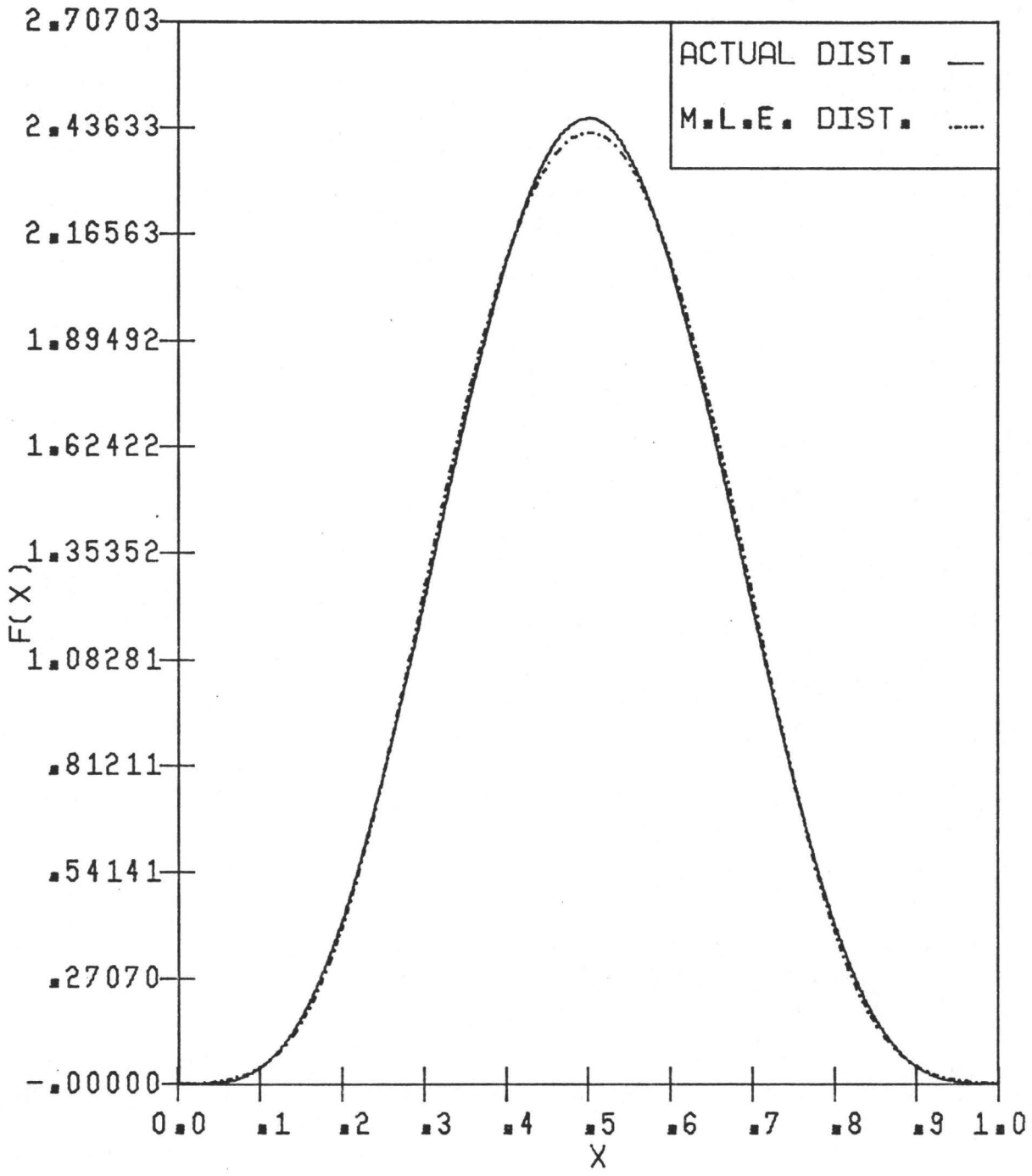


Figure A.8 Approximating Beta Distribution (B5), (η = 5.0, γ = 5.0), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .5000 .02273 0.000 .00131 0.0000

λ values (for M.L.E. Distribution): - 7.08586 + 50.40342 - 125.22834 + 151.87114 - 80.08444 3.12476

Percentage area in common between the two distributions = 99.26

BETA DISTRIBUTION (B6)

$$f(x) = \begin{cases} \frac{\Gamma(\eta + \gamma)}{\Gamma(\eta)\ \ (\gamma)} x^{\gamma - 1}(1 - x)^{\eta - 1}, & \eta = 3.0,\ \gamma = 1.5,\ 0 \leq x \leq 1 \\ 0, \text{ elsehwere} \end{cases}$$

Central moment values: .33351 .04037 .00414

.00414 .00108 .00673

Standardized moment measures ($\sqrt{\beta_1}$, β_2): .511 2.539

Upper and lower bounds: 1.00 0.00

Type of curve: (bell shaped)

Table A.10 Common area between Beta Distribution (B6), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	96.12
4	98.11
5	98.20
6	98.86

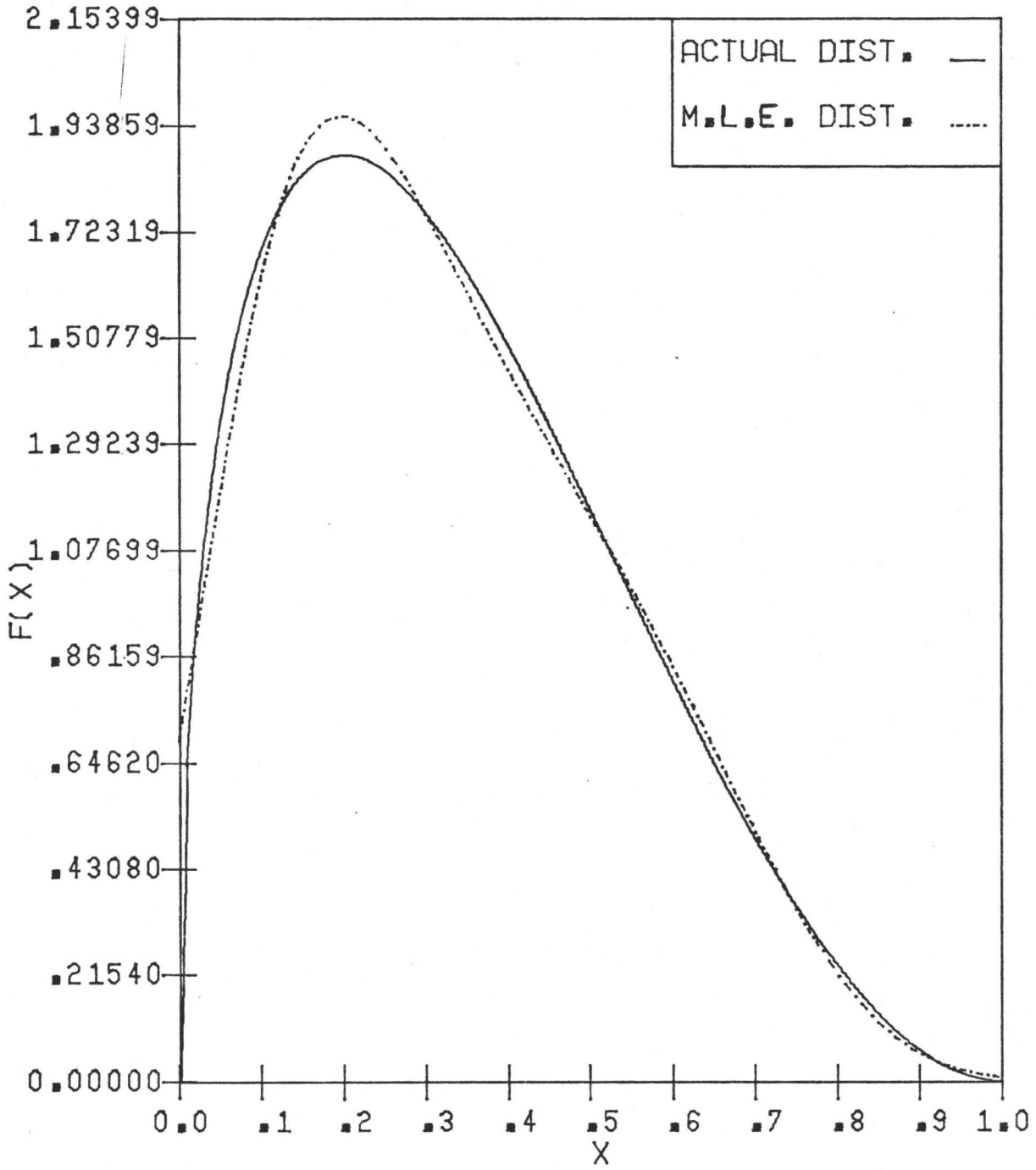


Figure A.9 Approximating Beta Distribution (B6), (η = 3.0, γ = 1.5) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .33351 .04037 .00414 .00414 .00108

λ values (for M.L.E. Distribution): - .36963 + 13.65424 - 61.74094 116.8970 - 102.58016 + 29.30392

Percentage area in common between the two distributions = 98.20

BETA DISTRIBUTION (B7)

$$f(x) = \begin{cases} \dfrac{\Gamma(\eta + \gamma)}{\Gamma(\eta)\,\Gamma(\gamma)}\, x^{\gamma - 1}(1 - x)^{\eta - 1}, & \eta = \gamma = 3.0,\ 0 \leqslant x \leqslant 1 \\ 0, \text{ elsewhere} & \end{cases}$$

Central moment values : .5000 .03571 0.000

.00298 0.000

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 0.00 2.333

Upper and lower bounds: 1.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.11 Common area between Beta Distribution (B7), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	95.54
4	98.57
5	98.56

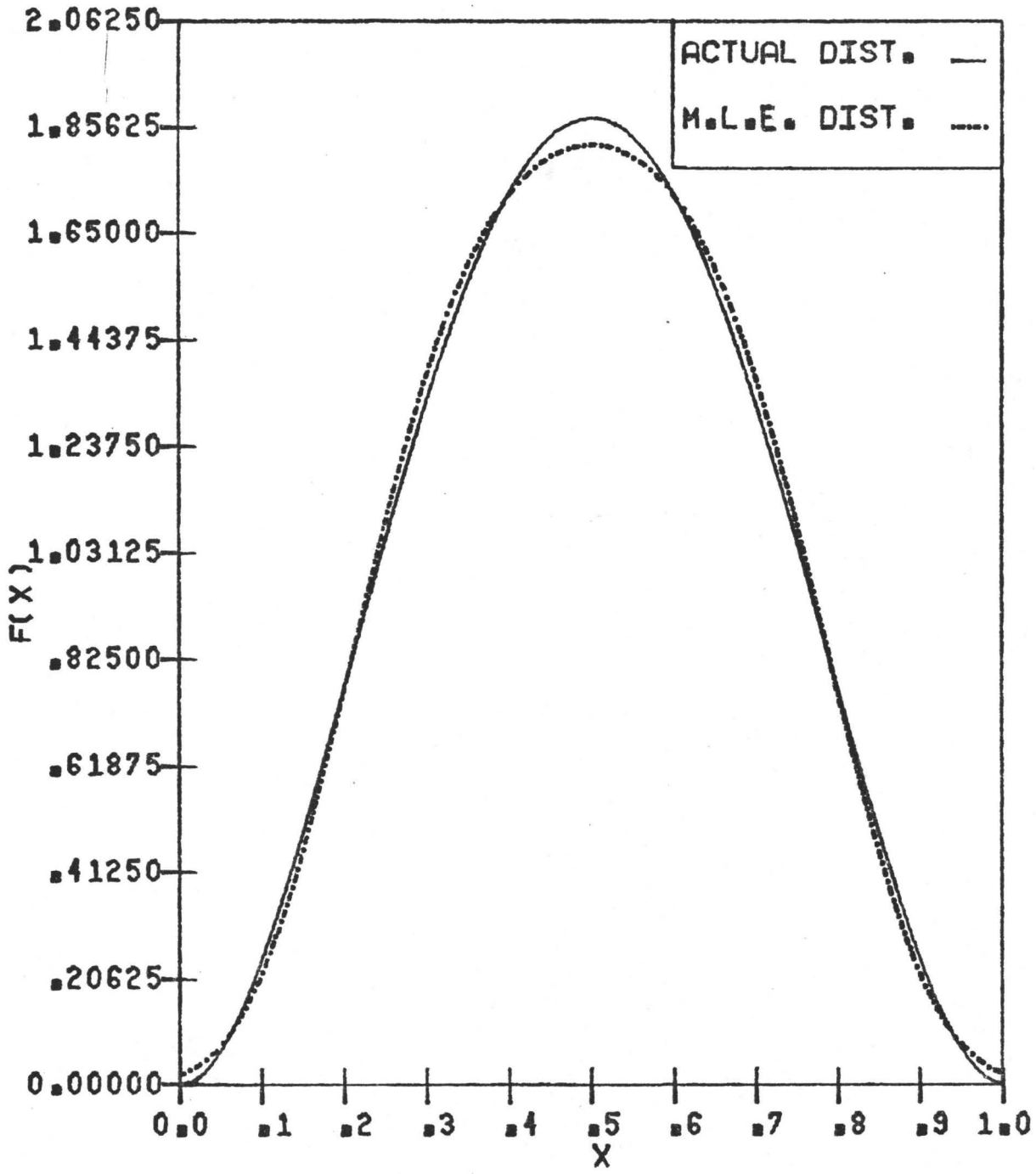


Figure A.10 Approximating Beta Distribution (B7), (η = 3.0, γ = 3.0) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .5000 .035714 0.000 .002976 0.0000

λ values (for M.L.E. Distribution): 3.94075 31.4819 - 85.0532 107.98137 - 55.5248 1.13750

Percentage area in common between the two distributions = 98.56

BETA DISTRIBUTION (B8)

$$f(x) = \begin{cases} \dfrac{\Gamma(\eta + \gamma)}{\Gamma(\eta)\,\Gamma(\gamma)}\, x^{\gamma - 1}(1 - x)^{\eta - 1}, & \eta = \gamma = 2.0,\ 0 \leqslant x \leqslant 1 \\ 0, \text{ elsewhere} & \end{cases}$$

Central moment values: .5000 .0500 0.0000

.005357 0.0000

Standarized moment measures ($\sqrt{\beta_1}$, β_2): 0.00 2.143

Upper and lower bounds: 1.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.12 Common area between Beta Distribution (B8), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments

Number of known first moments	Percentage of area in common between the two curves
3	95.57
4	98.17
5	98.17

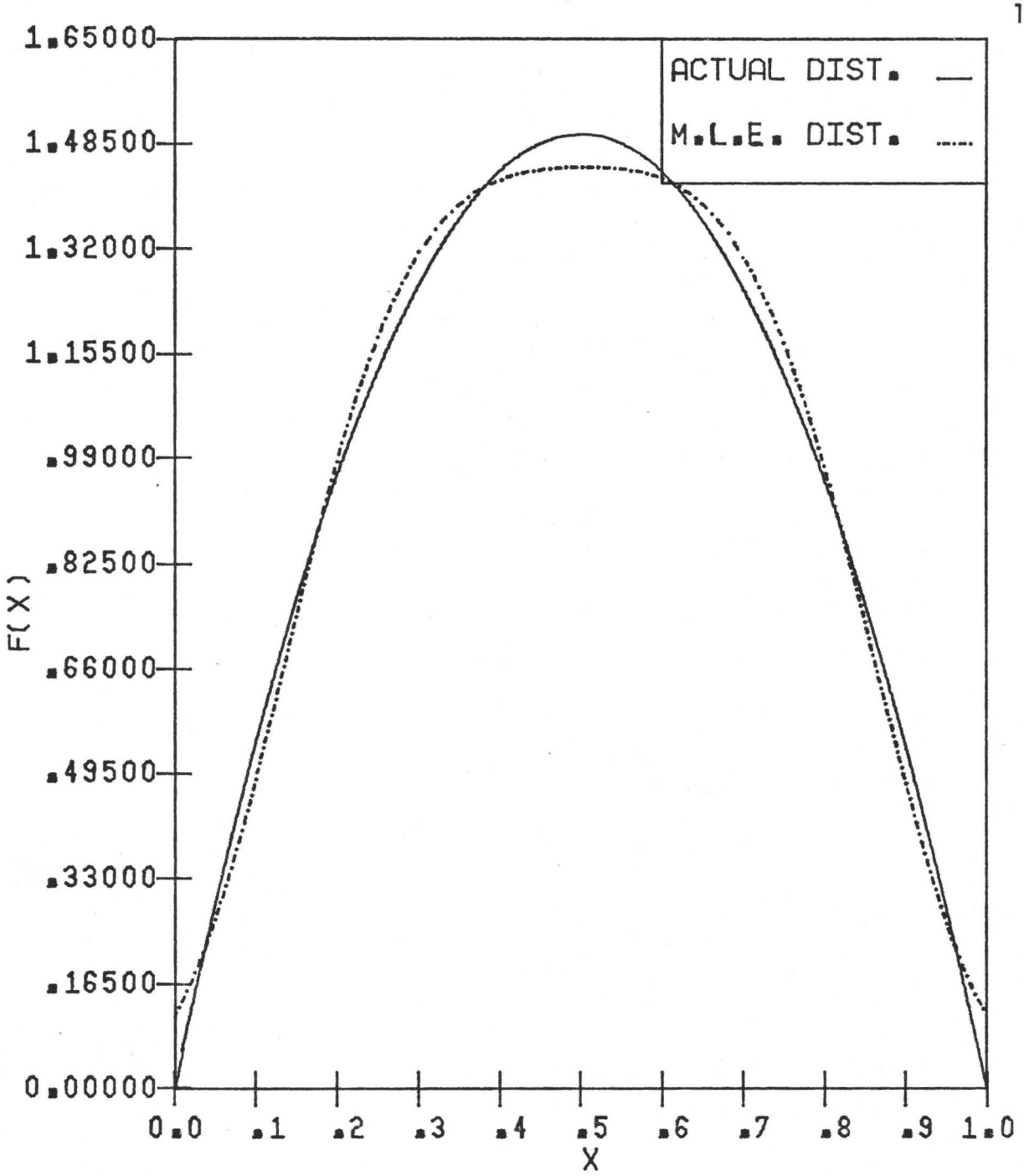


Figure A.11 Approximating Beta Distribution (B8), ($\eta = \gamma = 2.0$) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central moment values: .5000 .0500 0.0000 .00536 0.0000

λ values (for M.L.E. Distribution): - 2.1754 + 19.3931 - 56.47123 74.81196 - 38.6007 .87103

Percentage area in common between the two distributions = 98.17

BETA DISTRIBUTION (B9)

$$f(x) = \begin{cases} \dfrac{\Gamma(\eta + \gamma)}{\Gamma(\eta)\,\Gamma(\gamma)}\, x^{\gamma - 1}(1 - x)^{\eta - 1}, & \eta = 1.5,\ \gamma = 3.0,\ 0 \leqslant x \leqslant 1 \\ 0, \text{ elsewhere} & \end{cases}$$

Central moment values: .66649 .04037 -.00415

.004137 -.00108

Standardized moment measures ($\sqrt{\beta_1}$, β_2): -.511 +2.539

Upper and lower bounds: 1.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.13 Common area between Beta Distribution (B9), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	96.12
4	98.12
5	98.20

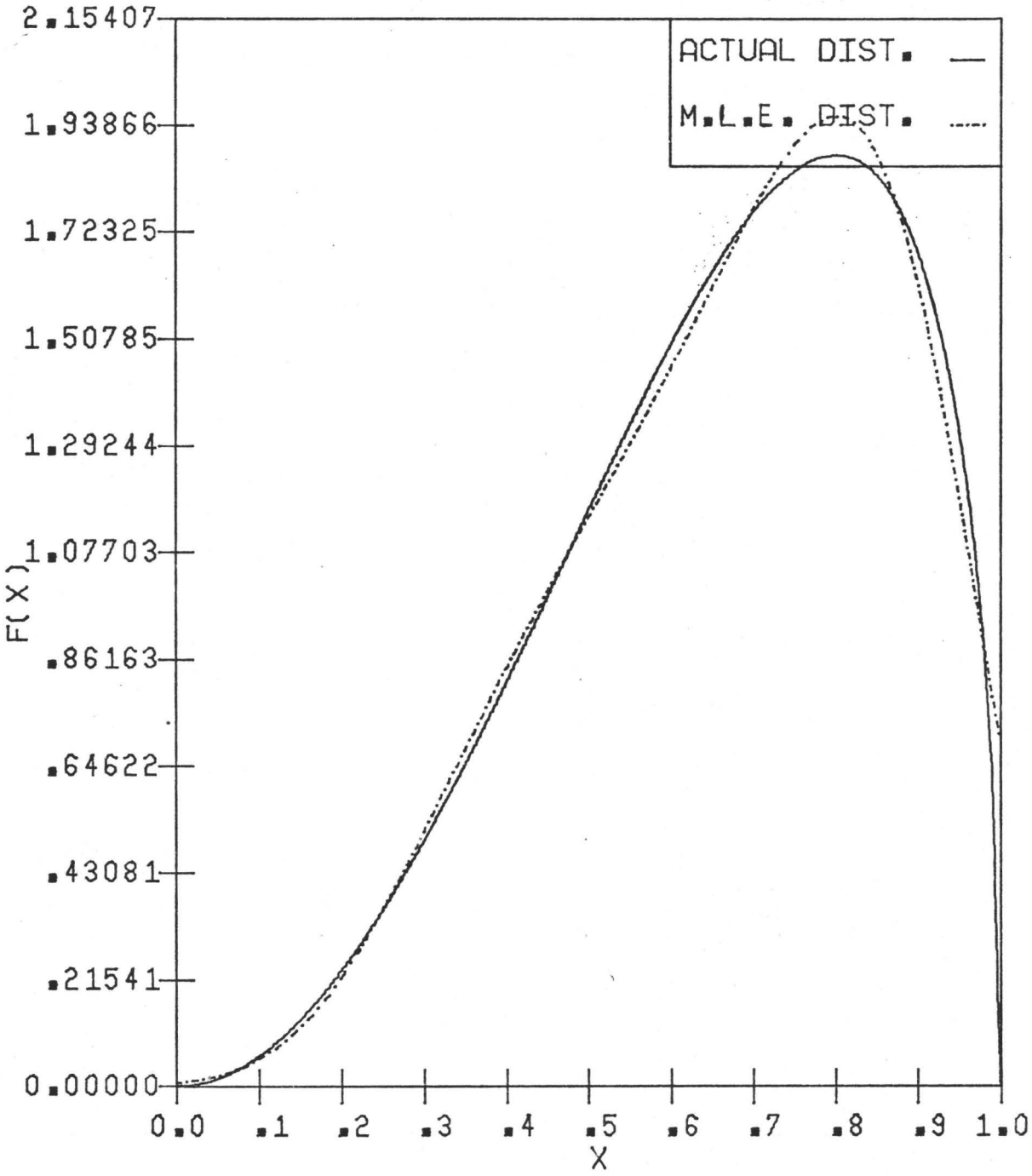


Figure A.12 Approximating Beta Distribution (B9), (η = 1.5, γ = 3.0) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: + .66649 + .04037 - .00415 + .004137 - .00108

λ values (for M.L.E. Distribution): - 4.9173 + 24.1267 - 39.1723 12.29497 32.57655 - 25.26896

Percentage area in common between the two distributions = 98.20

BETA DISTRIBUTION (B10)

$$f(x) = \begin{cases} \dfrac{\Gamma(\eta + \gamma)}{\Gamma(\eta)\,\Gamma(\gamma)}\, x^{\gamma - 1}(1 - x)^{\eta - 1}, & \eta = 1.5,\ \gamma = 5.0,\ 0 \leq x \leq 1 \\ 0, \text{ elsewhere} & \end{cases}$$

Central moment values: +.768973 +.0236348 -.002997
+.001832

Standardized moment measures ($\sqrt{\beta_1}$, β_2): -.825 +3.280

Upper and lower bounds: 1.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.14 Common area between Beta Distribution (B10), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	96.09
4	97.56

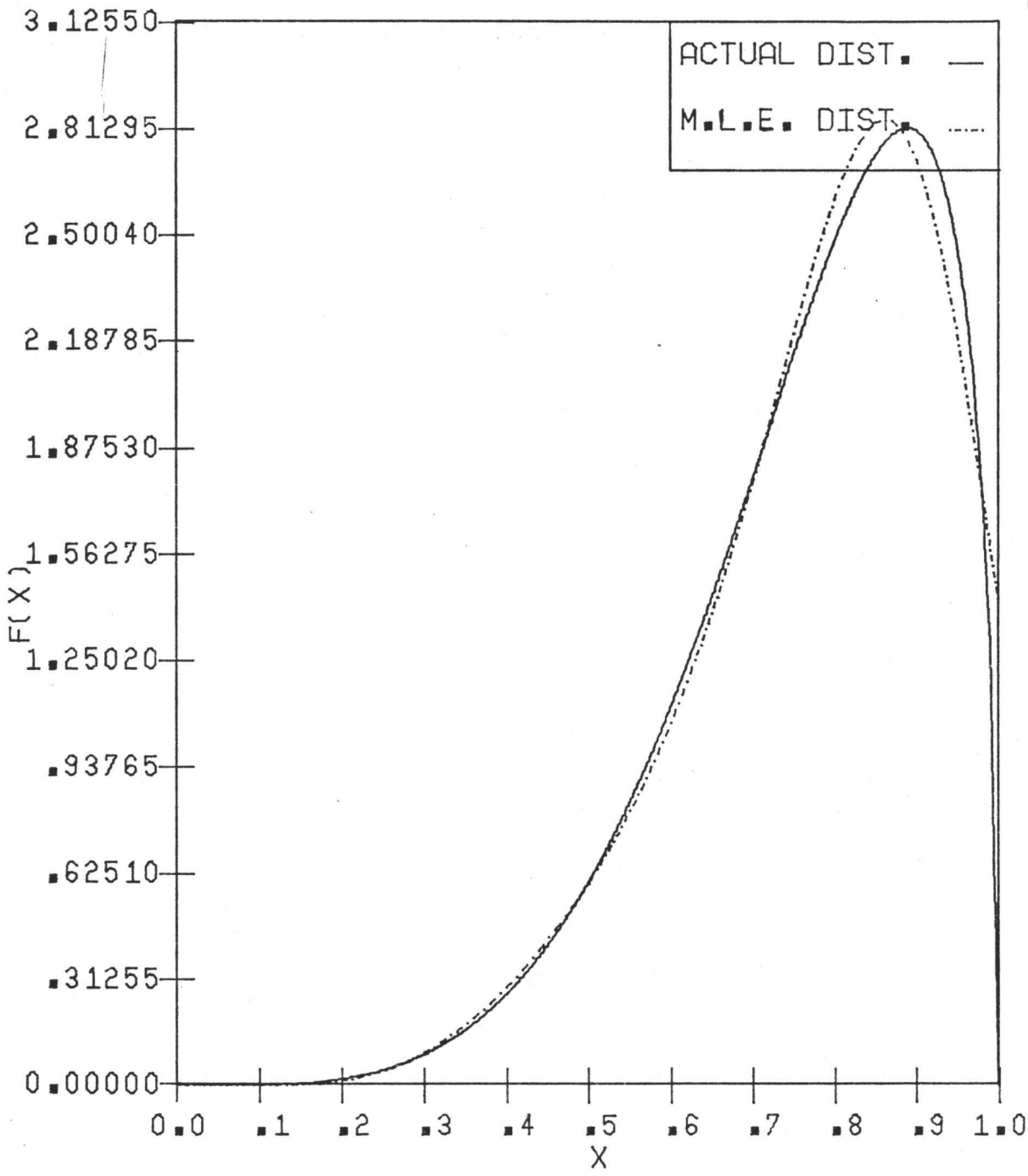


Figure A.13 Approximating Beta Distribution (B10), (η = 1.5, γ = 5.0) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first four moments.

Central Moment values: + .768973 + .0236348 - .002998 + .001832

λ values (for M.L.E. Distribution): -12.4111 63.21888 - 138.01770 .149.03429 - 61.478266

Percentage area in common between the two distributions = 97.56

BETA DISTRIBUTION (B11)

$$f(x) = \begin{cases} \frac{\Gamma(\eta + \gamma)}{\Gamma(\eta)\,\Gamma(\gamma)}\, x^{\gamma - 1}(1 - x)^{\eta - 1}, & \eta = 2.0,\ \gamma = .5,\ 0 \leq x \leq 1 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: .20524 .04585 .011998

.007889 .004053

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 1.222 3.753

Upper and lower bounds: 1.00 0.00

Type of curve: J shape

Table A.15 Common area between Beta Distribution (B11), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	90.42
4	91.01
5	92.50

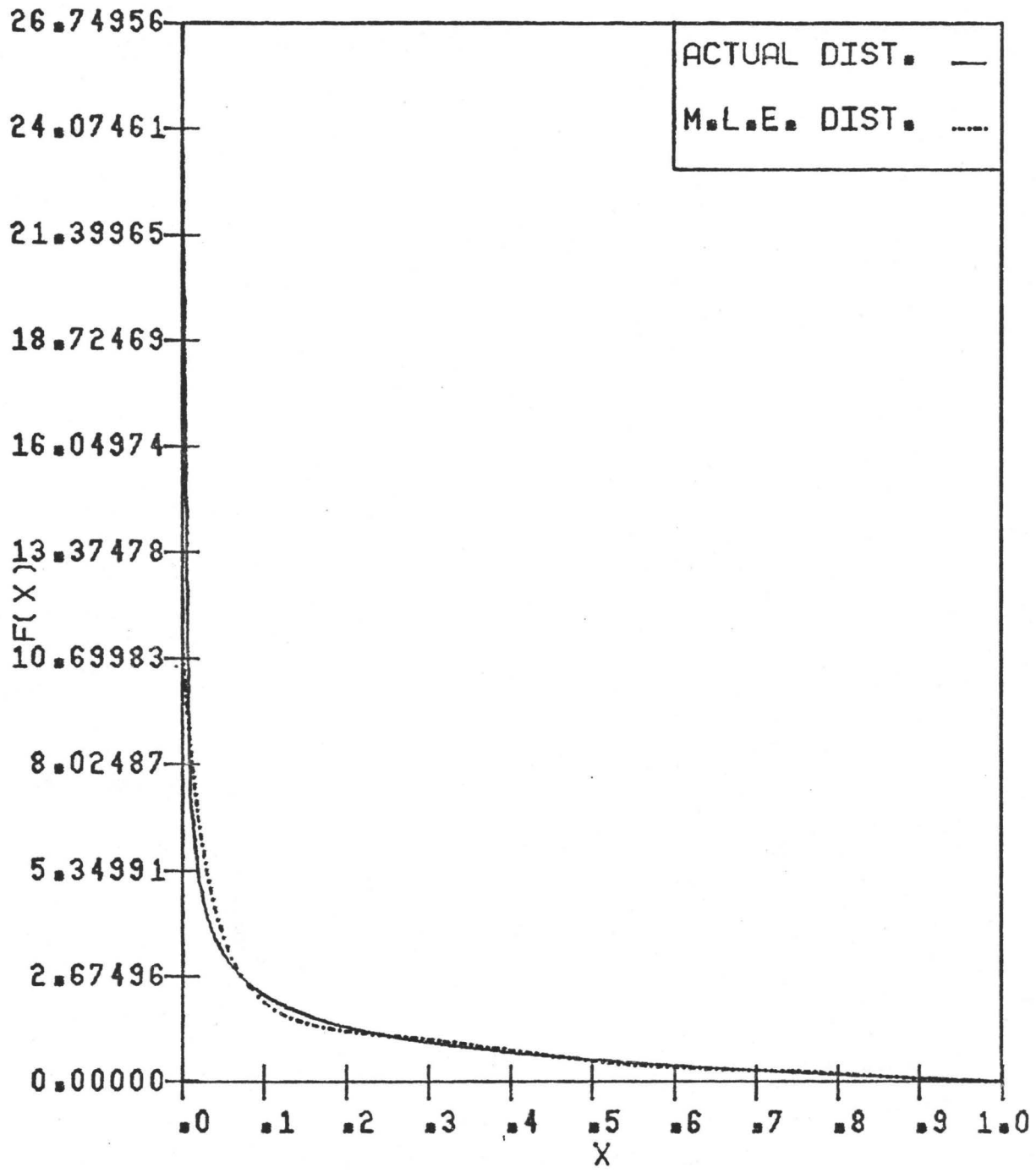


Figure A.14 Approximating Beta Distribution (B11), (η = 2.0, γ = .5) by a Maximum Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .20524 .04585 .011998 .007889 .00405

λ values (for M.L.E. Distribution): 2.44524 - 28.88730 148.03821 - 370.85069 418.5973 - 174.77754

Percentage area in common between the two distributions = 92.50

WEIBULL DISTRIBUTION (W1)

$$f(x) = \begin{cases} \frac{\eta}{\sigma}\left(\frac{x}{\sigma}\right)^{\eta - 1} \exp\left[-\left(\frac{x}{\sigma}\right)^{\eta}\right], & \eta = \sigma = 1.0,\ x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: .92631 .69574 .73785
2.0312 4.3413 10.9381

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 1.271 4.196

Upper and lower bounds: 4.00 0.00

Type of curve: J shape (exponential case)

Table A.16 Common area between Weibull Distribution (W1), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	99.99
4	99.98
5	99.96
6	99.95

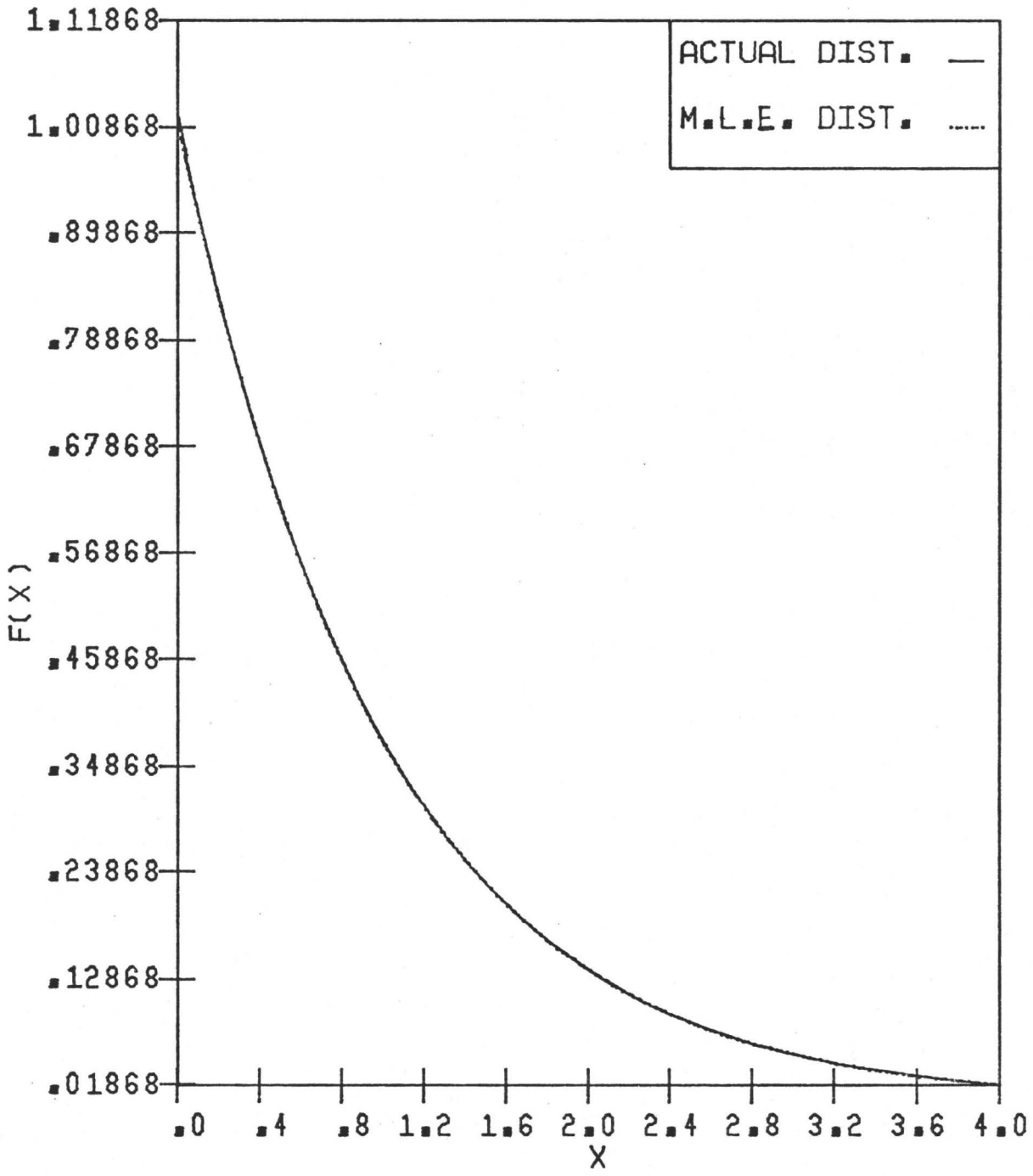


Figure A.15 Approximating Weibull Distribution (W1), ($\eta = \sigma = 1.0$) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .92631 .69574 .73785 2.0312 4.3413

λ values (for M.L.E. Distribution): + .017289 - .980712 - .038074 .027957 - .0085024 .0009065

Percentage area in common between the two distributions = 99.96

WEIBULL DISTRIBUTION (W2)

$$f(x) = \begin{cases} \frac{\eta}{\sigma}\left(\frac{x}{\sigma}\right)^{\eta-1} \exp\left[-\left(\frac{x}{\sigma}\right)^{\eta}\right], \eta = 4., \sigma = 1., x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: .90640 .06466 -.00143
.01149 -.00059

Standardized moment measures ($\sqrt{\beta_1}$, β_2): -.0872 2.7478

Upper and lower bounds: 3.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.17 Common area between Weibull Distribution (W2), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	98.24
4	99.38
5	99.66

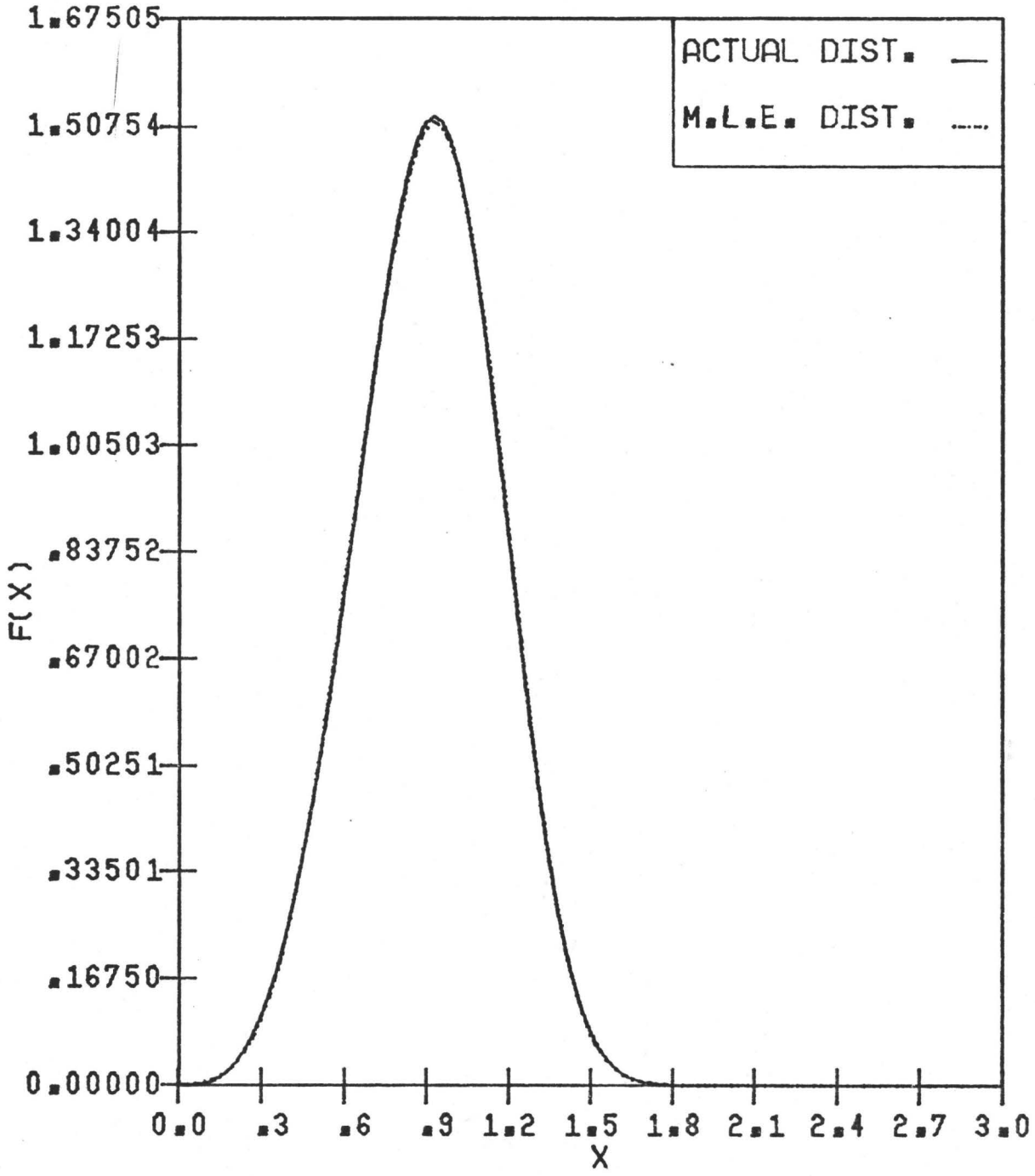


Figure A.16 Approximating Weibull Distribution (W2), (η = 4.0, σ = 1.0) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .90640 .06466 - .00143 .01149 - .00059

λ values (for M.L.E. Distribution): - 7.36687 25.7626 - 38.16523 33.78615 - 16.4386 2.80577

Percentage area in common between the two distributions = 99.66

WEIBULL DISTRIBUTION (W4)

$$f(x) = \begin{cases} \frac{\eta}{\sigma}\left(\frac{x}{\sigma}\right)^{\eta-1} \exp -\left(\frac{x}{\sigma}\right)^{\eta}, & \eta = 0.5,\ \gamma = 1.0,\ x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: .68306 .86862 1.35043

3.77670 9.73903 27.05002

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 1.668 5.006

Upper and lower bounds: 4.00 0.00

Type of curve: J shape

Table A.18 Common area between Weibull Distribution (W4), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
2	74.29
3	75.56
4	77.61
5	79.92
6	82.02

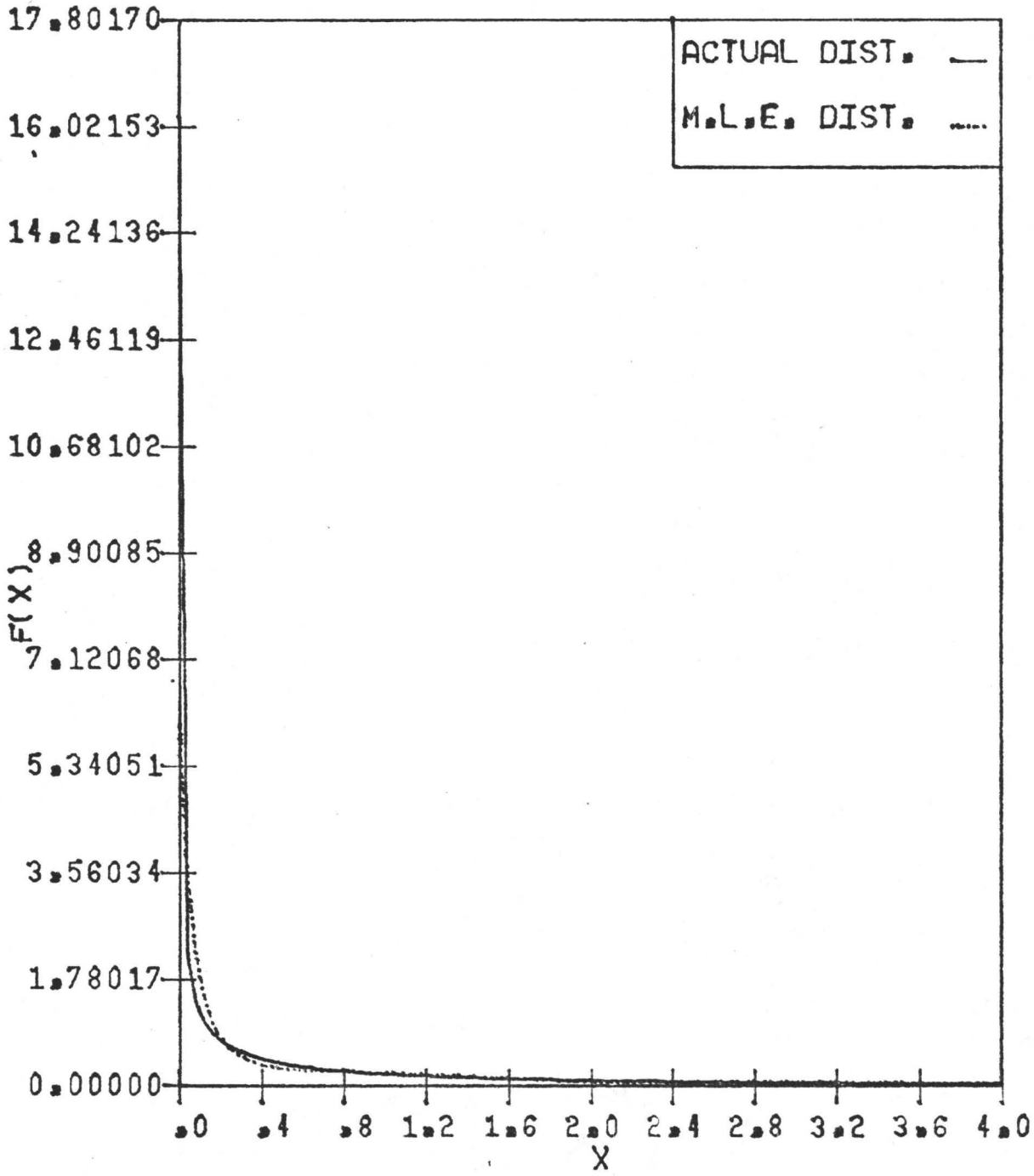


Figure A.17 Approximating Weibull Distribution (W4), (η = .5, γ = 1.0) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first six moments.

Central Moment values: .68306 .86862 1.35043 3.77670 9.73903 27.0500

λ values (for M.L.E. Distribution): + 1.81709 - 14.31629 + 25.21242 - 21.85061 + 9.44108 - 1.97430 + .15866

Percentage area in common between the two distributions = 82.02

EXPONENTIAL DISTRIBUTION (E)

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \lambda = 1,\ x \geqslant 0 \\ 0. \text{ elsewhere} \end{cases}$$

Central moment values: 1.000 .9954 1.9546
8.5187 3.89138

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 1.968 8.596

Upper and lower bounds: 10.0 0.00

Type of curve: J shape

Table A.19 Common area between Exponential Distribution (E), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
1	99.99
2	99.97
3	99.93
4	99.86
5	99.74

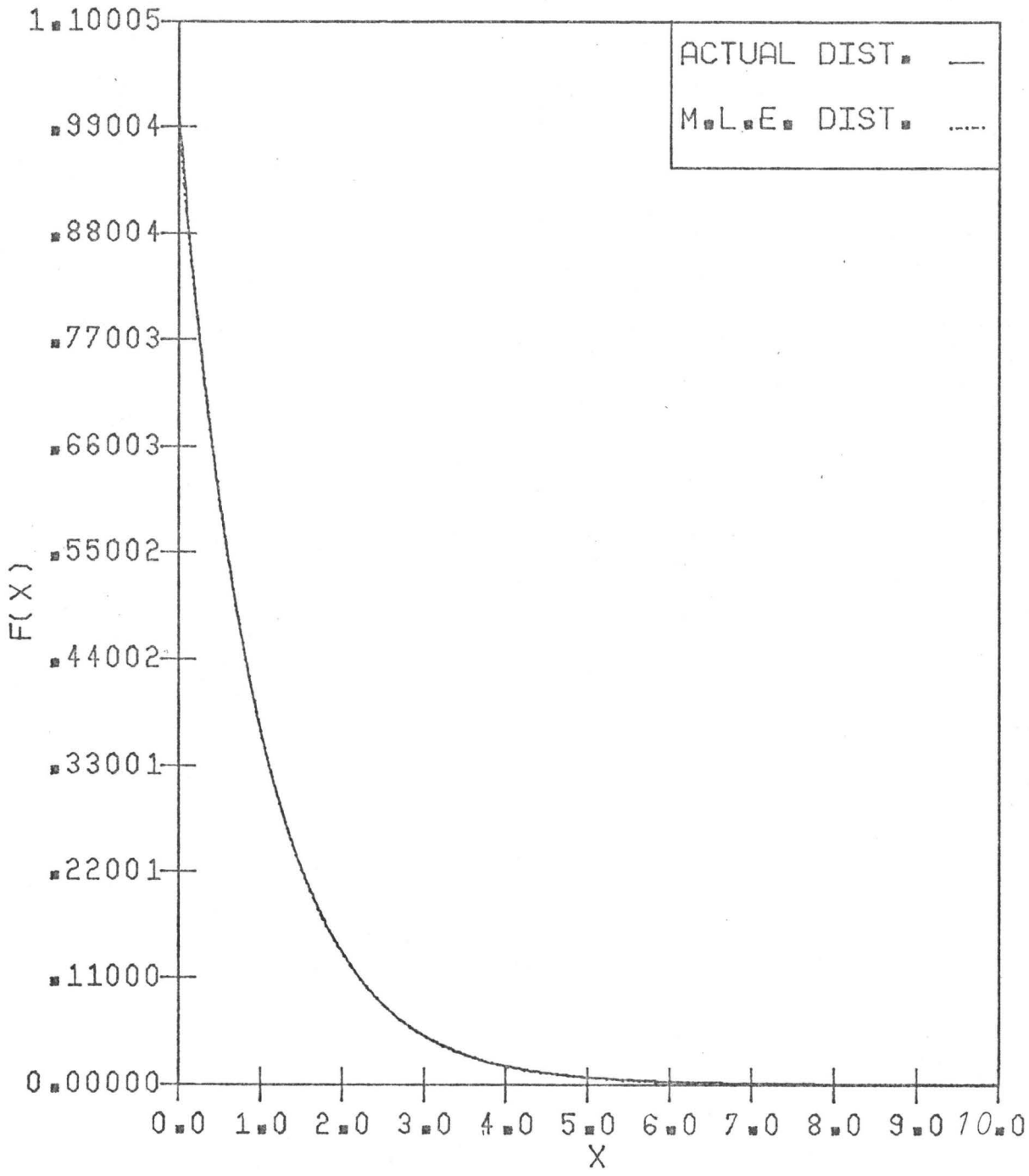


Figure A.18 Approximating Exponential Distribution (E), ($\lambda = 1$) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: 1.000 .9954 1.9546 8.5187 3.89138

λ values (for M.L.E. Distribution): - .01276 - .93661 - .06694 + .024997 - .003703 + .000185

Percentage area in common between the two distributions = 99.74

UNIFORM DISTRIBUTION (U)

$$f(x) = \begin{cases} \dfrac{1}{\mu_1 - \mu_0}, & \mu_0 = 0.0,\ \mu_1 = 1.0,\ \mu_0 \leq x \leq \mu_1 \\ \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: .500 .08333 0.000

.0125 0.000 .002232

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 0.00 1.80

Upper and lower bounds: 1.00 0.00

Type of curve: rectangular

Table A.20 Common area between Uniform Distribution (U), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
1	100.00
2	100.00
3	100.00
4	100.00
5	99.99
6	99.82

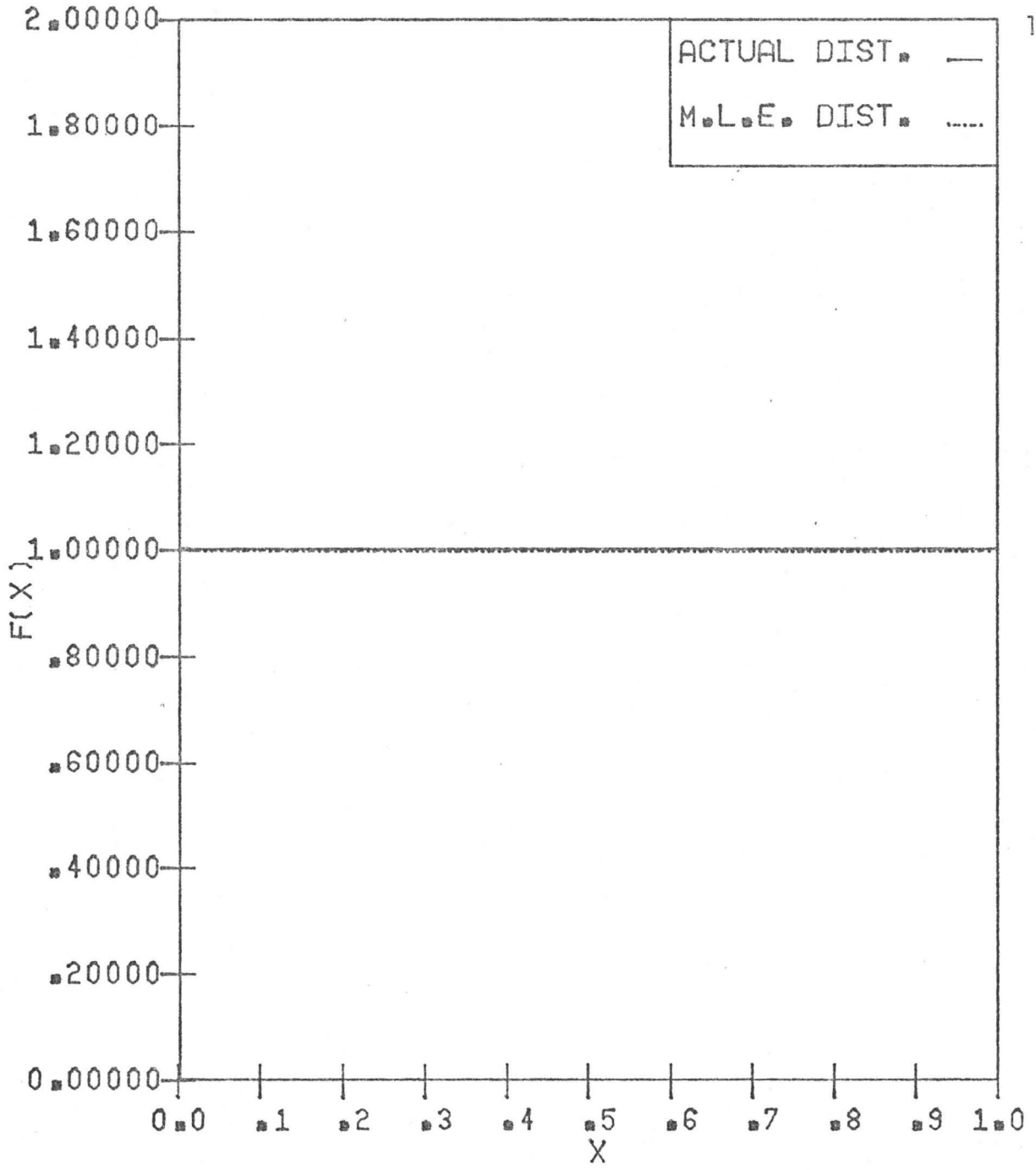


Figure A.19 Approximating Uniform Distribution (U), ($\mu_0 = 0$, $\mu_1 = 1$), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first moments.

Central Moment value: .500

λ values (for M.L.E. Distribution): 0.000 0.000

Percentage area in common between the two distributions = 100.00

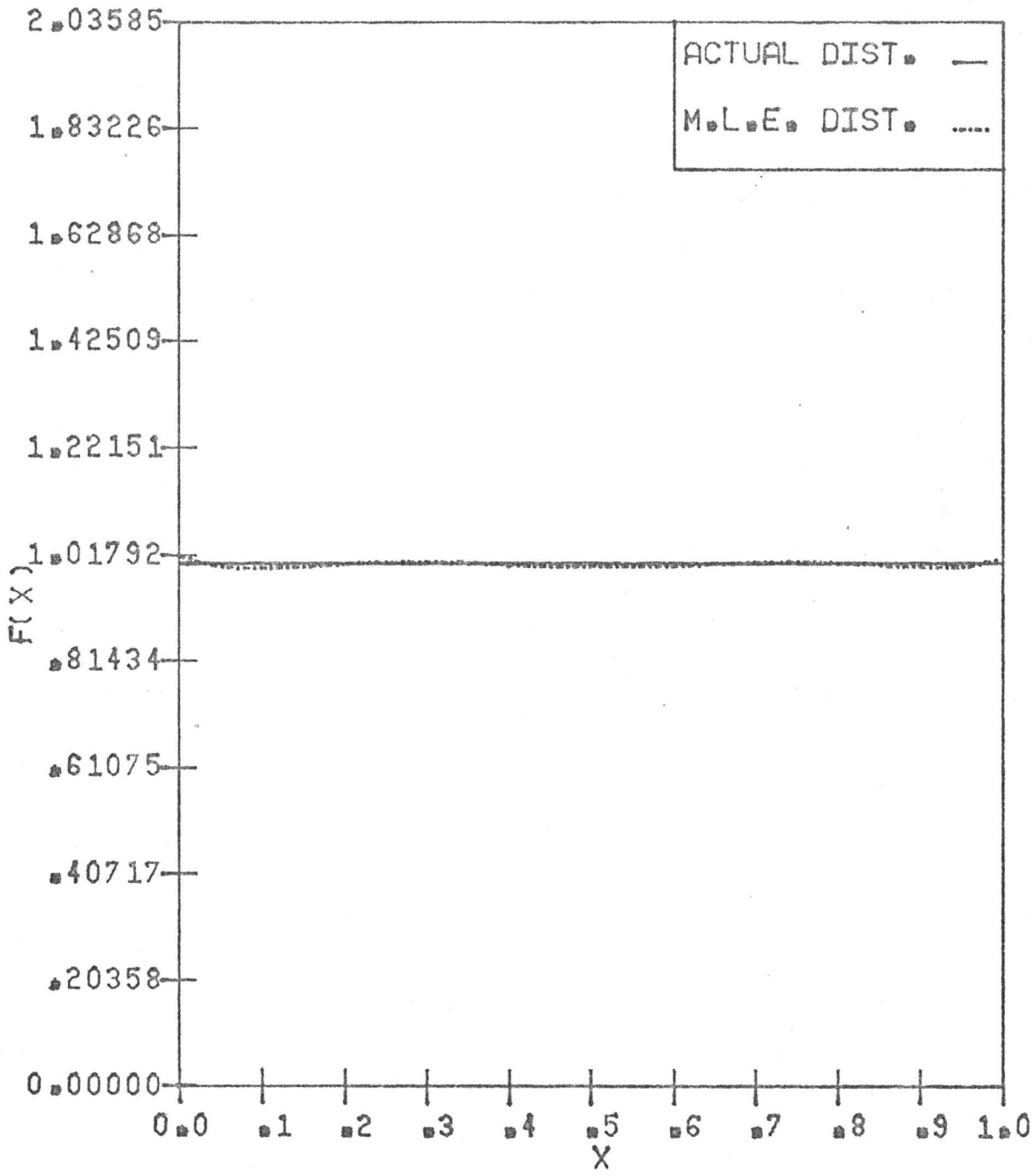


Figure A.20 Approximating Uniform Distribtuion (U), ($\mu_0 = 0$, $\mu_1 = 1$) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first six moments.

Central Moment values: .500 .08333 0.000 .01250 0.000 .002232

λ values (for M.L.E. Distribution): + .01777 - .71100 + 6.84518 - 26.5716 + 48.6262 - 41.9398 + 13.7454

Percentage area in common between the two distributions = 99.82

NORMAL DISTRIBUTION (N)

$$f(x) = \begin{cases} \frac{1}{\sigma\ 2\Pi} \exp\left[-(x - \mu)^2 / 2\sigma^2\right], & \mu = 4.0,\ \sigma = 1.0 \quad -\infty < x < +\infty \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 4.000 .99893 0.000

2.97966 0.000 14.62418

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 0.00 2.99

Upper and lower bounds: 8.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.21 Common area between Normal Distribution (N), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
2	100.00
3	100.00
4	99.99
5	99.99
6	100.00

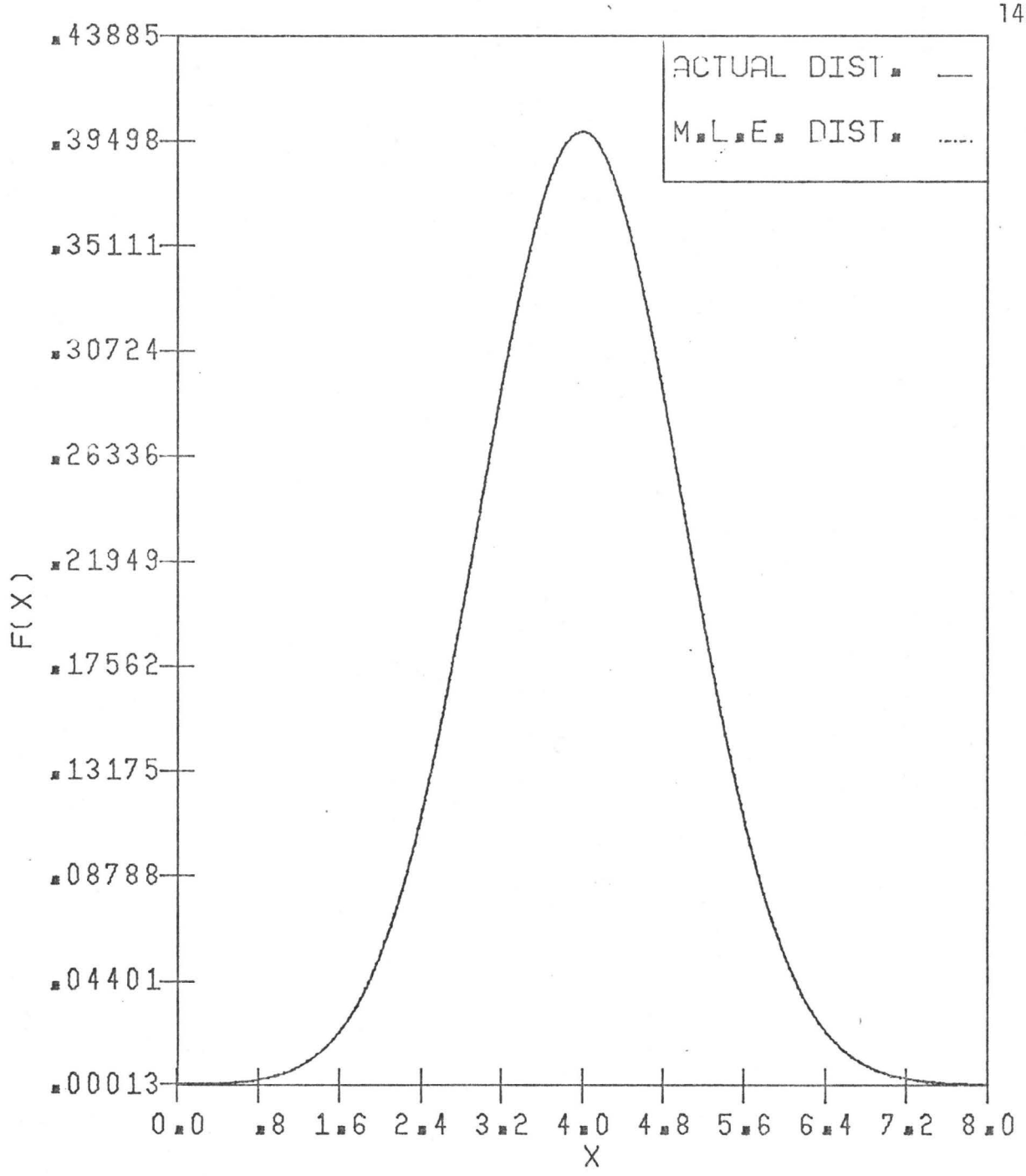


Figure A.21 Approximating Normal Distribution (N), ($\mu = 4.0$, $\sigma = 1.0$) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first two moments.

Central Moment values: 4.00 .99893

λ values (for M.L.E. Distribution): - 8.91870 3.99991 - 4.99989

Percentage area in common between the two distributions = 100.00

TRUNCATED NORMAL DISTRIBUTION (TN)

$$f(x) = \begin{cases} \frac{1}{\sigma\ 2\Pi} \exp\left[-(x - \mu)^2/2\sigma^2\right], & \mu = 2.0,\ \sigma = 2.0,\ -\infty < x < \infty \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 2.0000 1.16450 0.00000

2.63201 0.00000 7.27213

Standardized moment measures($\sqrt{\beta_1}$, β_2): 0.00 1.941

Upper and lower bounds: 4.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.22 Common area between Truncated Normal Distribution (TN), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
2	100.00
3	100.00
4	100.00
5	100.00
6	99.89

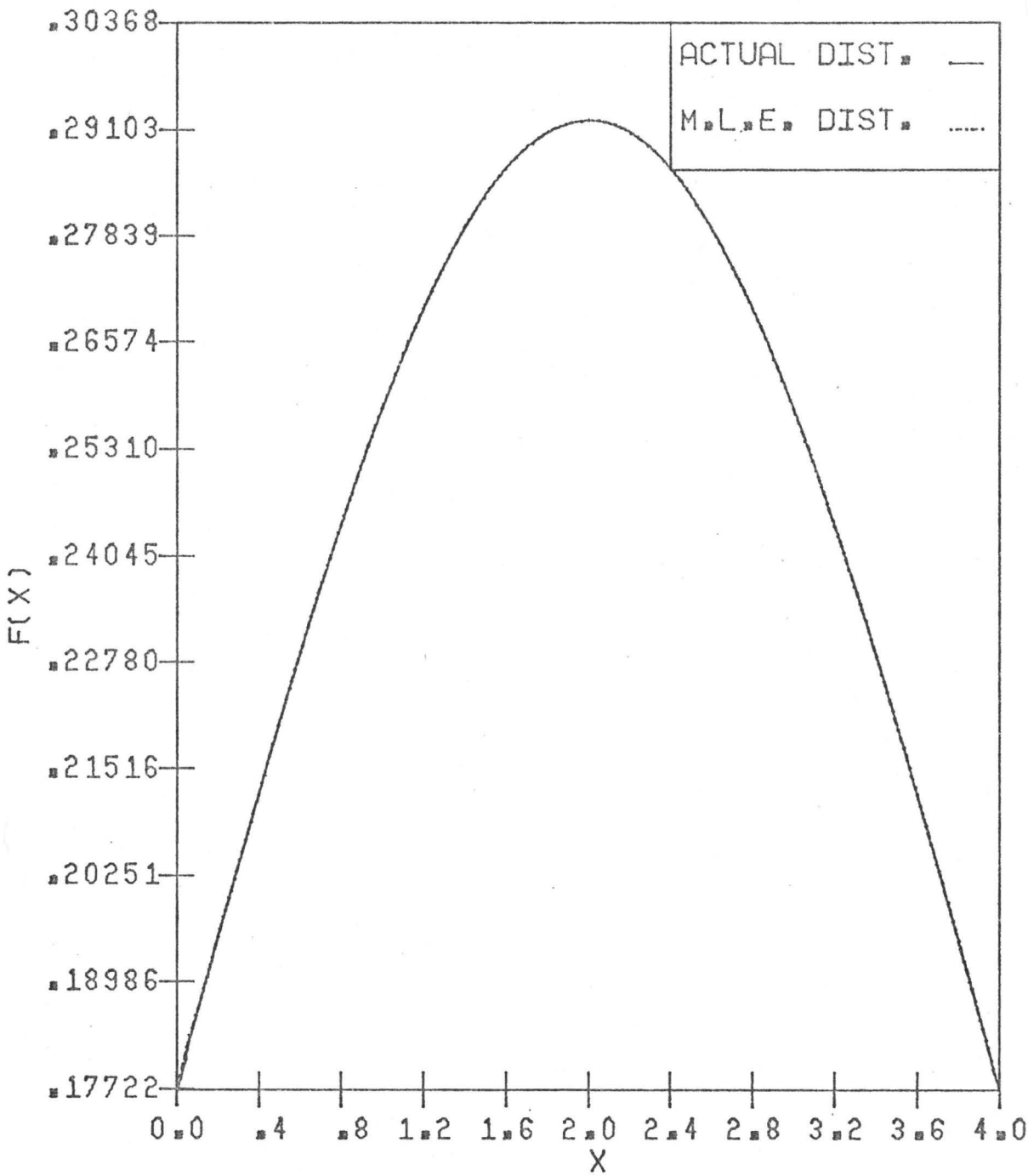


Figure A.22 Approximating Truncated Normal Distribution (TN), (μ = 2.0, σ = 2.0) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first two moments.

Central Moment values: 2.0000 1.16450

λ values (for M.L.E. Distribution): - 1.730363 + 0.49999 - .124998

Percentage area in common between the two distributions = 100.00

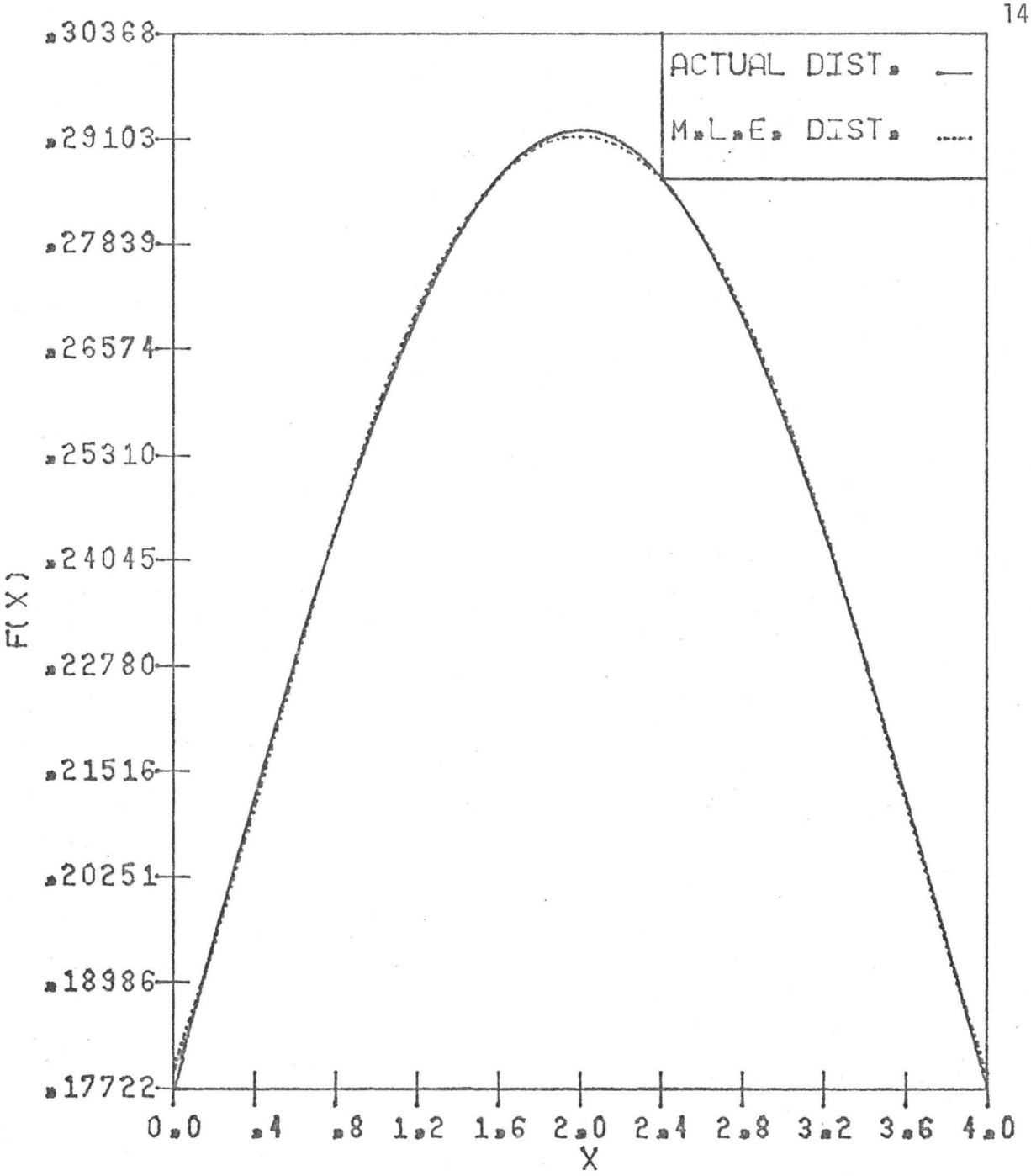


Figure A.23 Approximating Truncated Normal Distribution (TN), ($\mu = 2.0$, $\sigma = 2.0$) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first six moments.

Central Moment values: 2.000 1.1645 0.000 2.63201 0.000 7.27213

λ values (for M.L.E. Distribution): - 1.71775 + .38162 + .15274 - .26399 + .11989 - .02578 + .002112

Percentage area in common between the two distributions = 99.89

HALF NORMAL DISTRIBUTION (HN)

$$f(x) = \begin{cases} \left(\frac{2}{\Pi\sigma^2}\right)^{1/2} \exp\left[-x^2/2\sigma^2\right], \sigma = 5.0, x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 3.989411 9.084231 27.2459

319.1838 2458.863

Standardized moment measures ($\sqrt{\beta_1}$, β_2): .995 3.868

Upper and lower bounds: 25.00 0.00

Type of curve: J shape

Table A.23 Common area between Half Normal Distribution (HN), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
2	100.00
3	100.00
4	100.00
5	99.98

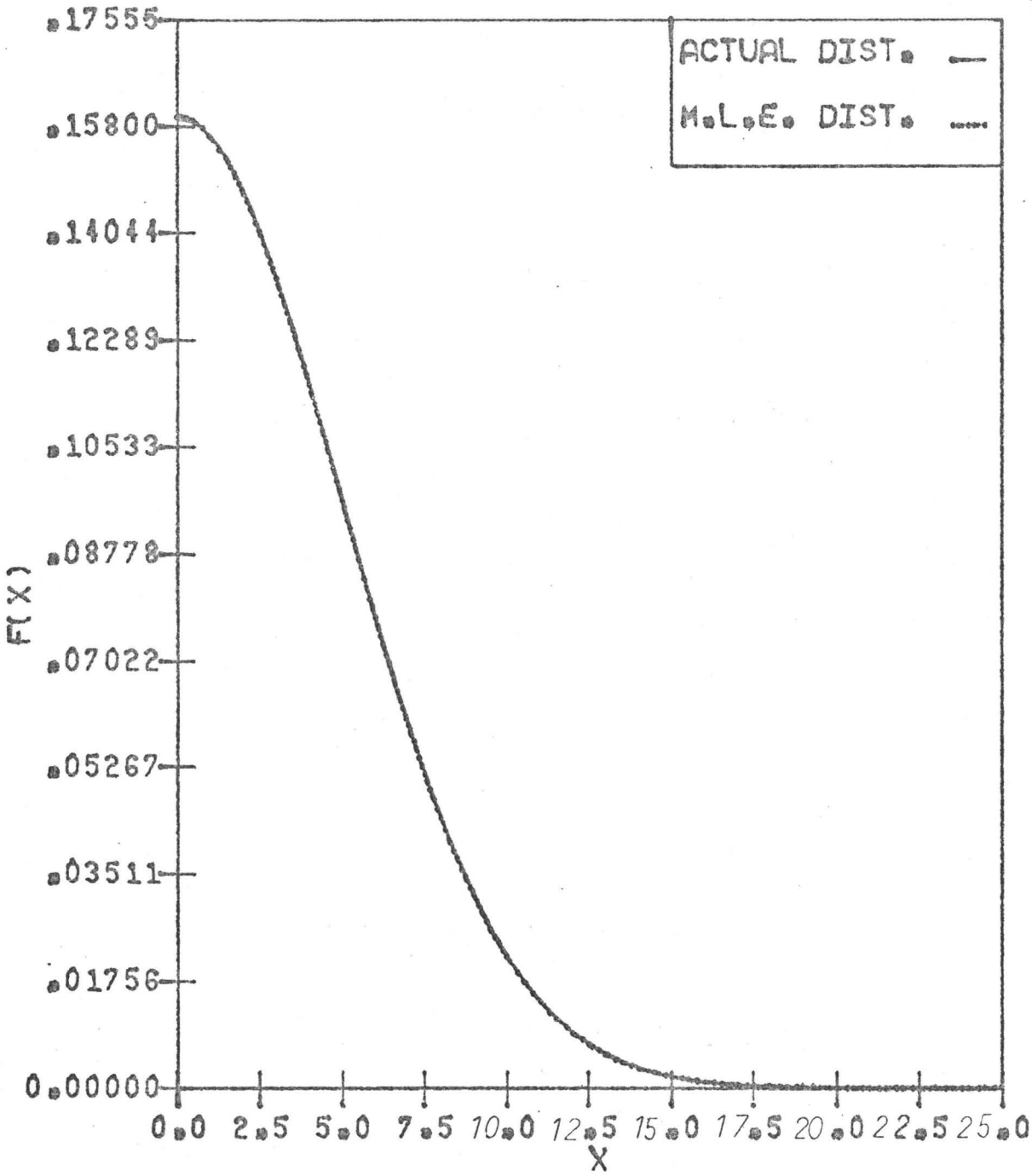


Figure A.24 Approximating Half Normal Distribution (HN), (σ = 5.0) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first two moments.

Central Moment values: 3.9894 9.0842

λ values (for M.L.E. Distribution): - 1.835137 - .00004 - .019996

Percentage area in common between the two distributions = 100.00

RAYLEIGH DISTRIBUTION (R1)

$$f(x) = \begin{cases} \left(\frac{x}{\sigma^2}\right) \exp\left(-x^2/2\sigma^2\right), \ \sigma^2 = 2.0, \ x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 1.77253 .85840 .501907
2.39098 4.09360

Standardized moment measures($\sqrt{\beta_1}$, β_2): .631 3.245

Upper and lower bounds: 8.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.24 Common area between Rayleigh Distribution (R1) and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	96.52
4	97.76
5	98.22

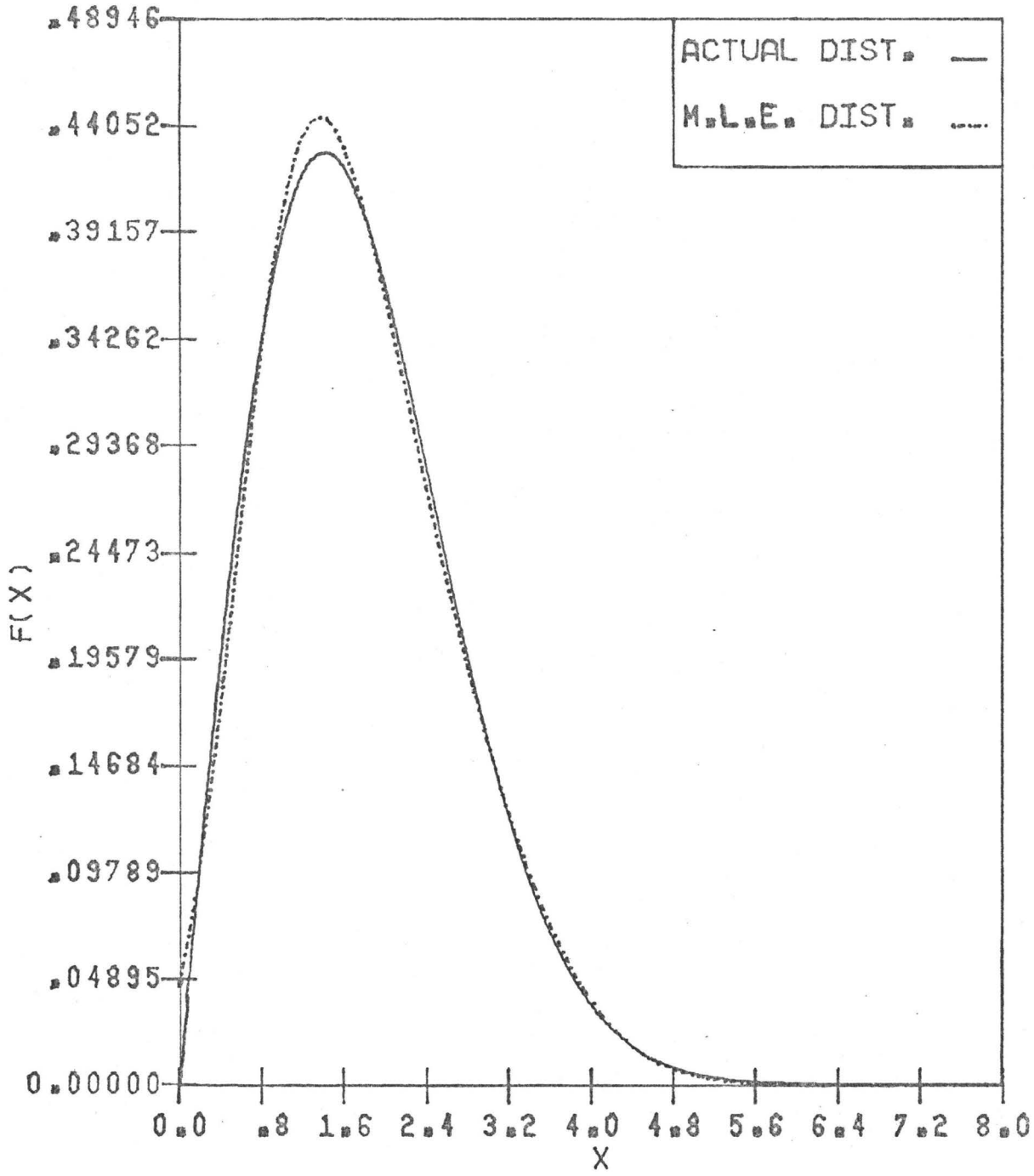


Figure A.25 Approximating Rayleigh Distribution (R1), (σ^2 = 2.) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: 1.77253 .85840 .501907 2.39098 4.09360

λ values (for M.L.E. Distribution): - 3.0918 4.3300 - 2.8808 - .84631 - .12897 .00730

Percentage area in common between the two distributions = 99.22

RAYLEIGH DISTRIBUTION (R2)

$$f(x) = \begin{cases} \left(\frac{x}{\sigma^2}\right) \exp\left(-x^2/2\sigma^2\right), \sigma^2 = 1, x \geqslant 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 1.25231 .42634 .16973
.571095 .64245 1.46602

Standardized moment measures ($\sqrt{\beta_1}$, β_2): .6097 3.1419

Upper and lower bounds: 4.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.25 Common area between Rayleigh Distribution (R2), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	96.68
4	97.82
5	98.45
6	98.86

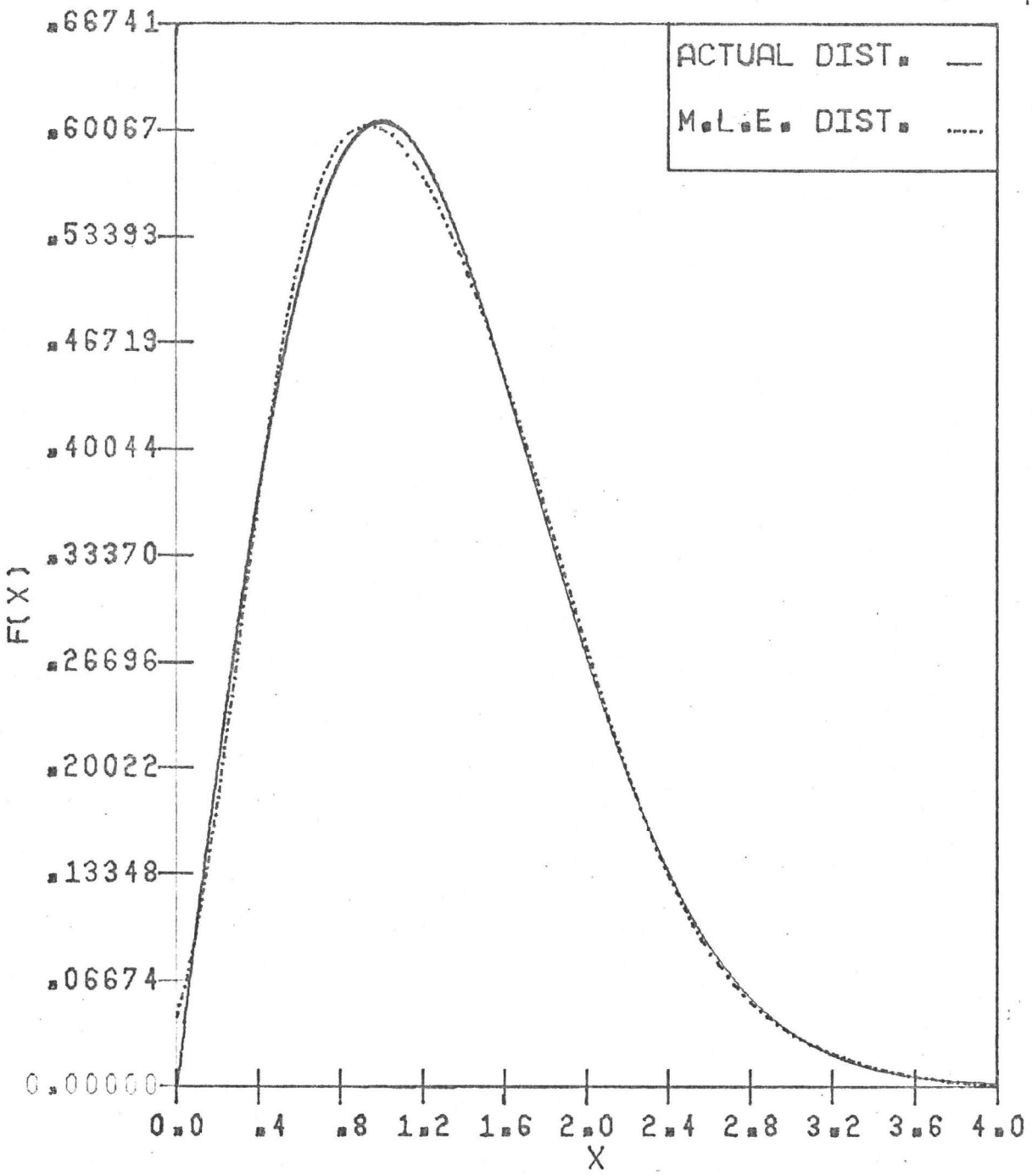


Figure A.26 Approximating Rayleigh Distribution (R2), (σ^2 = 1) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first six moments.

Central Moment values : 1.25231 .42634 .571095 .64245 1.46602

λ values (for M.L.E. Distribution): - 3.1339 9.1667 - 12.8105 9.4097 - 3.8796 .8051 - .06577

Percentage area in common between the two distributions = 98.86

RAYLEIGH DISTRIBUTION (R3)

$$f(x) = \begin{cases} \left(\frac{x}{\sigma^2}\right) \exp\left(-x^2/2\sigma^2\right), \ \sigma^2 = .5, \ x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: .88595 .21399 .06146
.14615 .12033

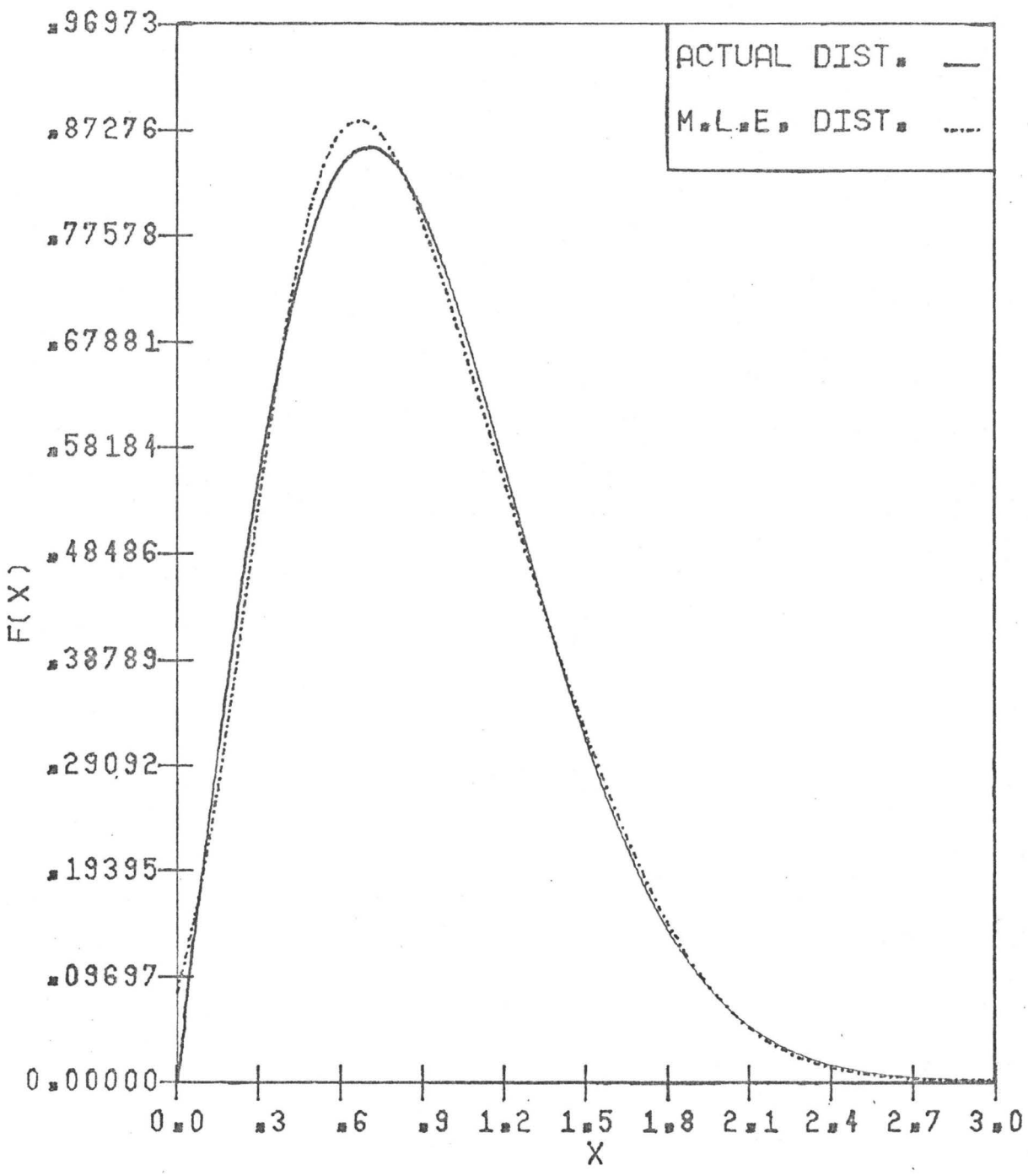
Standardized moment measures($\sqrt{\beta_1}$, β_2): .6209 3.1917

Upper and lower bounds: 3.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.26 Common area between Rayleigh Distribution (R3), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	96.63
4	97.79
5	98.41



FigureA.27 Approximating Rayleigh Distribution (R3), (σ^2 = .5) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: .88595 .21399 .06146 .14615 .12033

λ values (for M.L.E. Distribution): - 2.5017 + 9.6311 - 14.1161 9.5524 - 3.3372 .4398

Percentage area in common between the two distributions = 98.41

CAUCHY DISTRIBUTION (C1)

$$f(x) = \begin{cases} \frac{1}{\sigma \Pi} \left[1 + \frac{(x - \mu)^2}{\sigma^2}\right]^{-1}, & \mu = 0,\ \sigma = 1,\ -\infty < x < \infty \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 0.0000 44.8540 0.0000
151,158.9 0.0000

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 0.000 75.13

Upper and lower bounds: + 200.0 - 200.0

Type of curve: $\cap$ (bell shaped)

Table A.27 Common area between Cauchy Distribution (C1), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments.	Percentage of area in common between the two curves
1	13.01
2	45.78
3	45.78
4	46.69

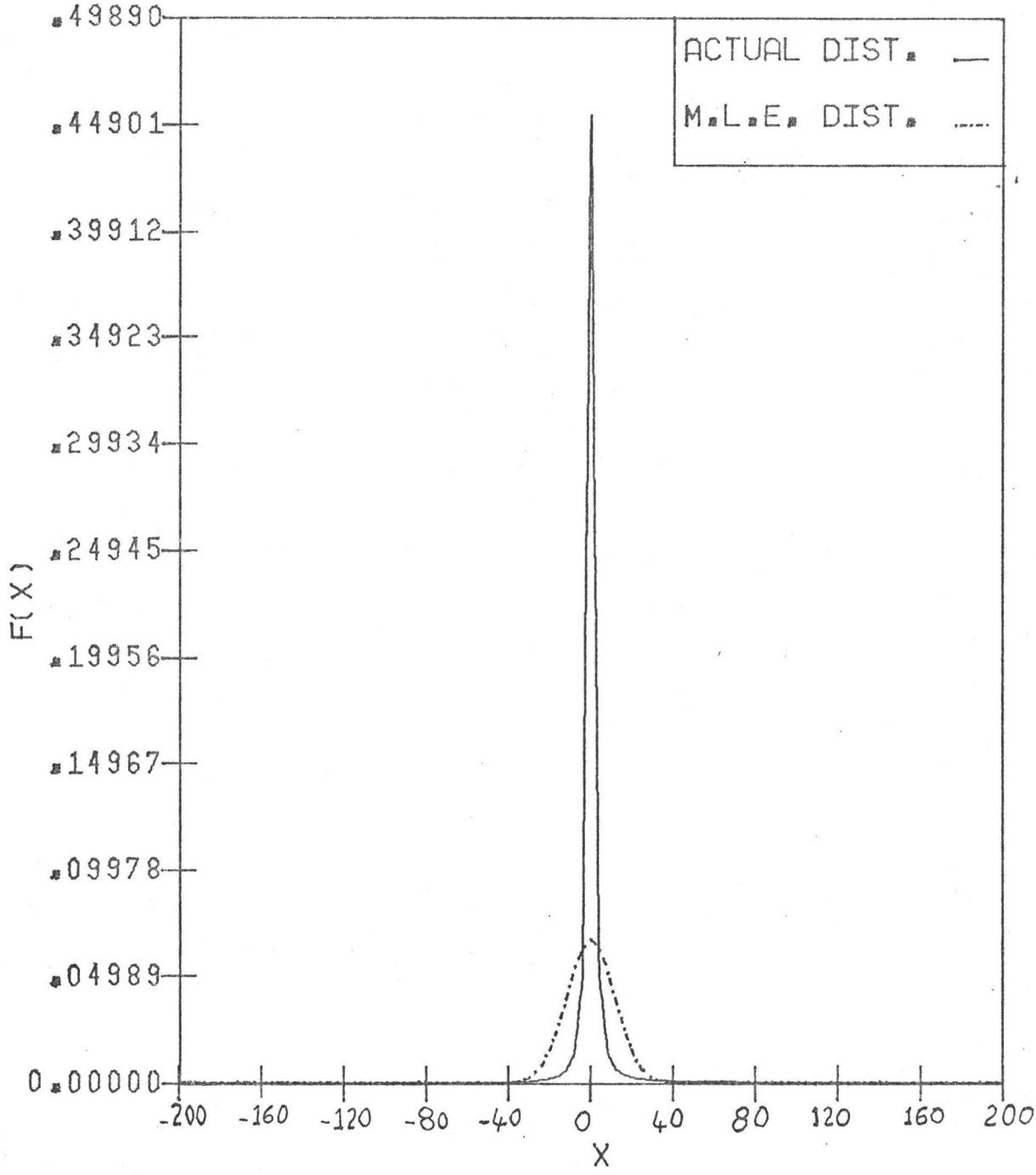


Figure A.28 Approximating Cauchy Distribution (C1), by a Maximum-Logarithmic Entropy Distribution (M.L.E.) based on the first four moments.

Central moment values: 0.000 44.8540 0.00000 151158.044

λ values (for M.L.E. Distribution): 2.69621 .00000 - .014968 .00000 0.00000

Percentage area in common between the two distributions = 46.69

CAUCHY DISTRIBUTION (C2)

$$f(x) = \begin{cases} \frac{1}{\sigma \Pi}\left[1 + \frac{(x - \mu)^2}{\sigma^2}\right]^{-1}, & \mu = 0.0,\ \sigma = 1.0,\ -\infty < x < \infty \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 0.0000 8.7345 0.0000

1,226.89 0.0000

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 0.00 16.08

Upper and lower bounds: 20.0 - 20.0

Type of curve: $\cap$ (bell shaped)

Table A.28 Common area between Cauchy Distribution (C2), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of knwon first moments	Percentage of area in common between the two curves
1	27.80
2	61.61
3	61.61
4	66.27
5	66.27

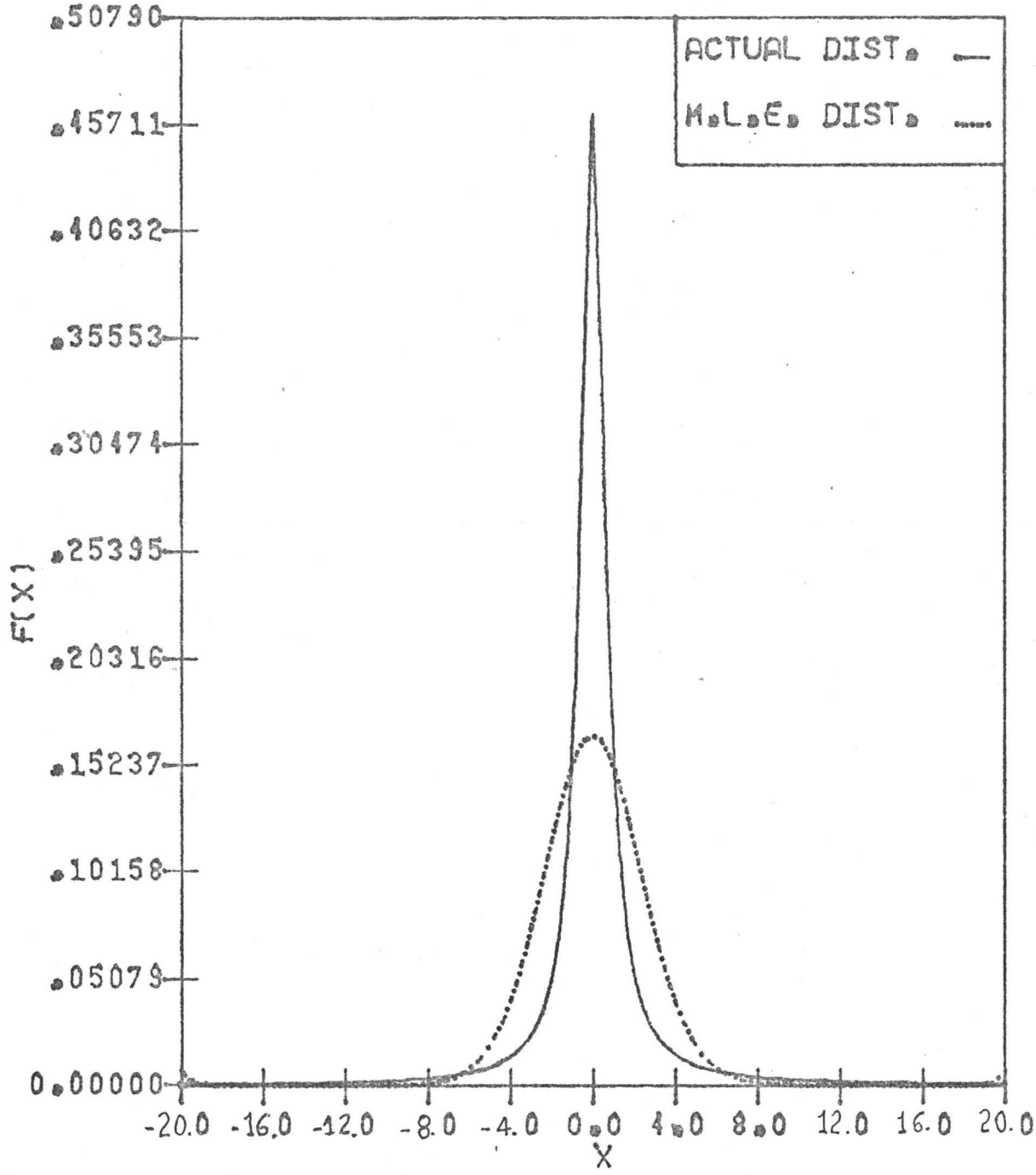


Figure A.29 Approximating Cauchy Distribution (C2) by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central moment values: 0.00 8.7345 0.00 1226.90 0.00

λ values (for M.L.E. Distribution): - 1.79216 - .001225 - .092266 .00007 .00021 0.00000

Percentage area in common between the two distributions = 66.27

LOG-NORMAL DISTRIBUTION (LN1)

$$f(x) = \begin{cases} \dfrac{1}{\sigma x\sqrt{2\Pi}} \exp\left[-\dfrac{1}{2\sigma^2}(\log x - \mu)^2\right], & \mu = 1,\ \sigma^2 = 1,\ x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 2.8239 2.39416 2.5889

15.5442 37.1409

Standardized moment measures ($\sqrt{\beta_1}$, β_2): .6989 2.7120

Upper and lower bounds: 7.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.29 Common area between Log-Normal Distribution (LN1), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	92.4
4	95.53
5	97.22

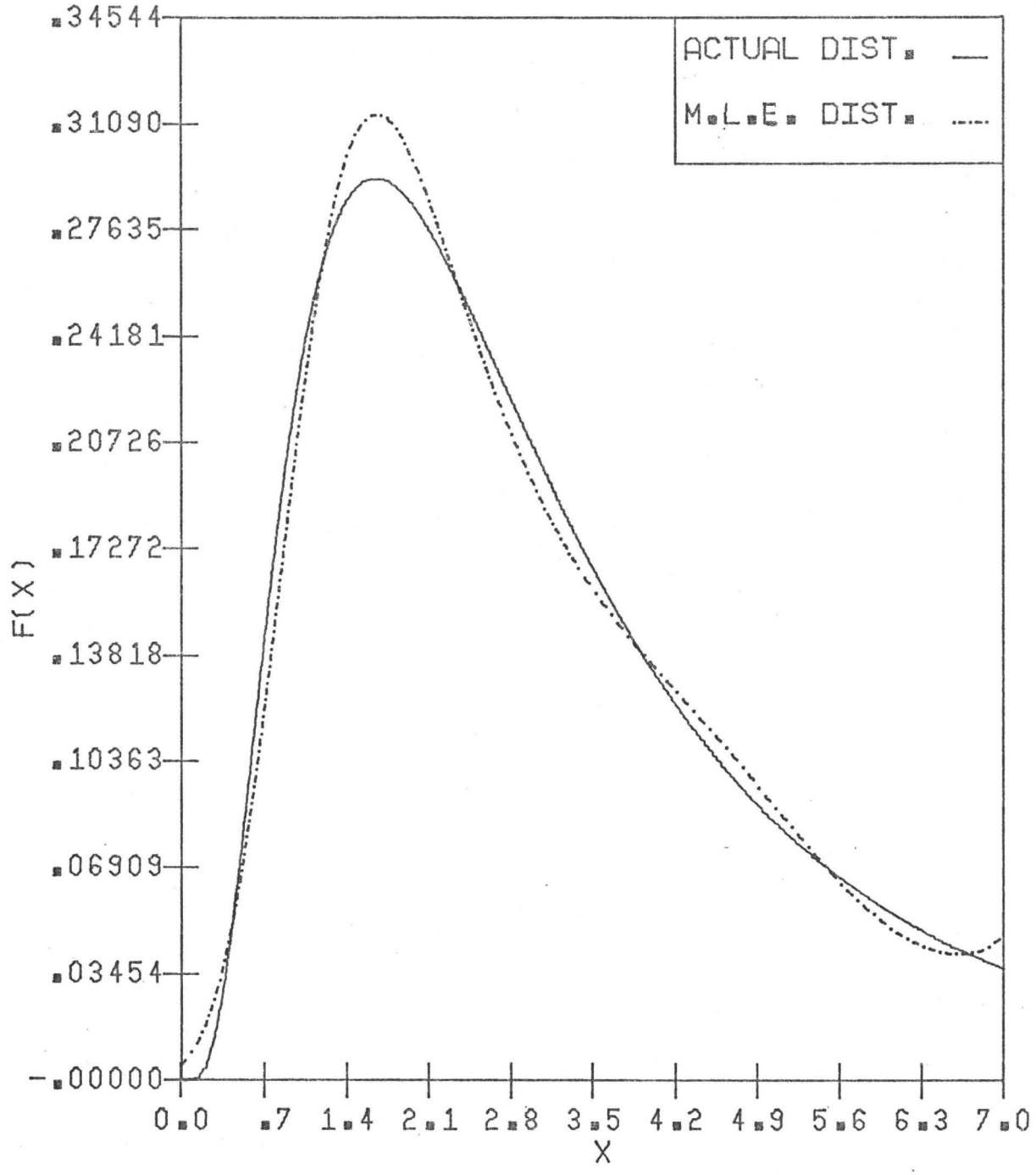


Figure A.30 Approximating Log-Normal Distribution (LN1), ($\mu = \sigma^2 = 1$), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: 2.8239 2.39416 2.58897 15.5442 37.1409

λ values (for M.L.E. Distribution): - 5.3522 7.0834 - 4.3202 1.18333 - .153852 .0076098

Percentage area in common between the two distributions = 97.22

LOG-NORMAL DISTRIBUTION (LN2)

$$f(x) = \begin{cases} \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right], & \mu = .3,\ \sigma^2 = 1.0,\ x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 1.6587 1.3486 2.4380

10.4366 38.5384

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 1.557 5.738

Upper and lower bounds: 7.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.30 Common area between Log-Normal Distribution (LN2), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	87.95
4	91.91
5	94.33

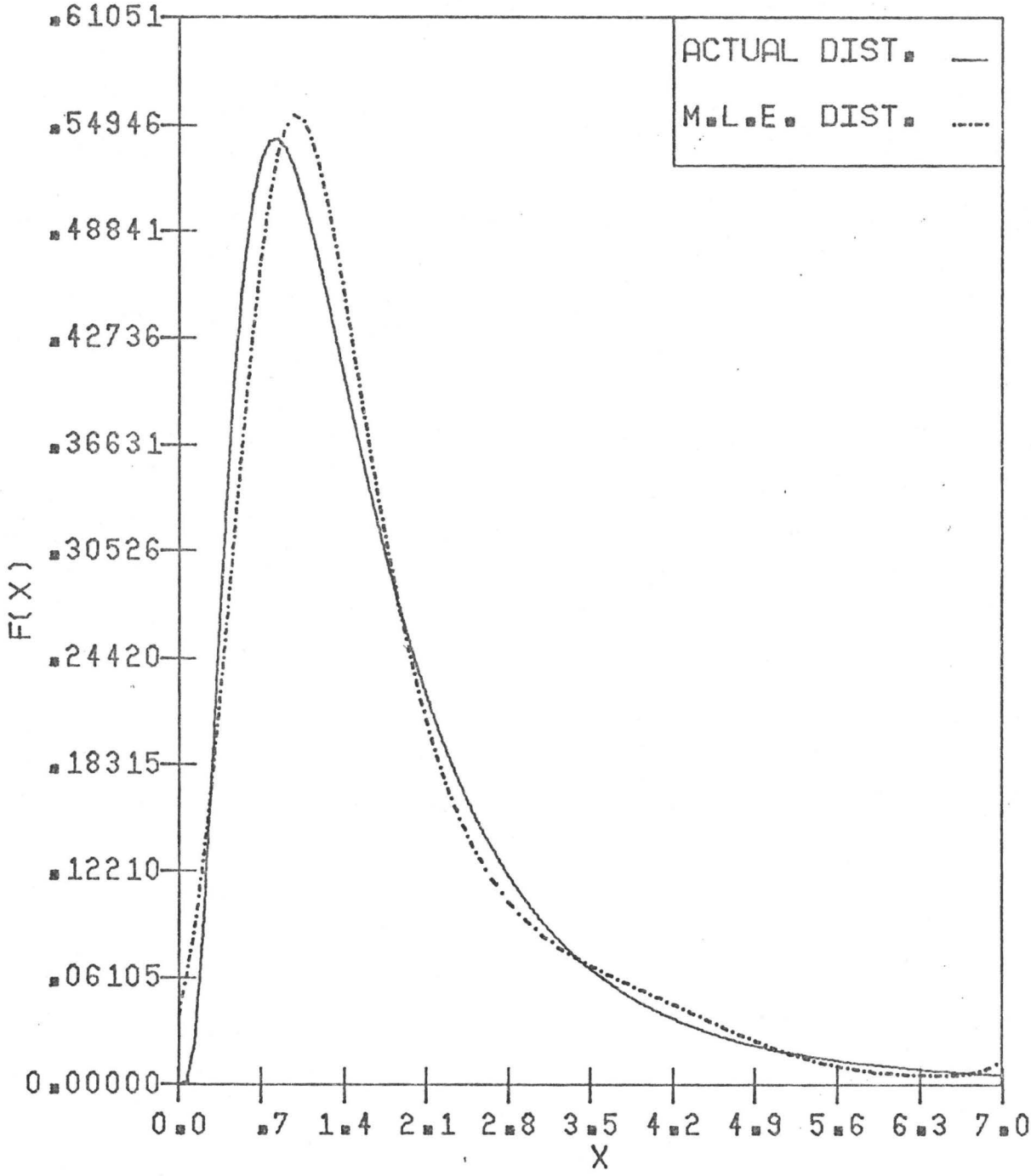


Figure A.31 Approximating Log-Normal Distribution (LN2), (μ = .3, σ^2 = 1.0), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: 1.65869 1.348597 2.437959 10.436564 38.538422

λ values (for M.L.E. Distribution): - 3.23115 6.62032 - 5.54092 1.8183437 - .269721 .01475896

Percentage area in common between the two distributions = 94.33

LOG-NORMAL DISTRIBUTION (LN3)

$$f(x) = \begin{cases} \frac{1}{\sigma \Pi \sqrt{2\Pi}} \exp\left[-\frac{1}{2\Pi^2}(\log x - \mu)^2\right], & \mu = 0,\ \sigma^2 = 1,\ x \geqslant 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 1.26162 .88814 1.6333

6.45511 24.81569

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 1.951 8.184

Upper and lower bounds: 7.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.31 Common area between Log-Normal Distribution (LN3), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	86.08
4	90.27
5	92.84

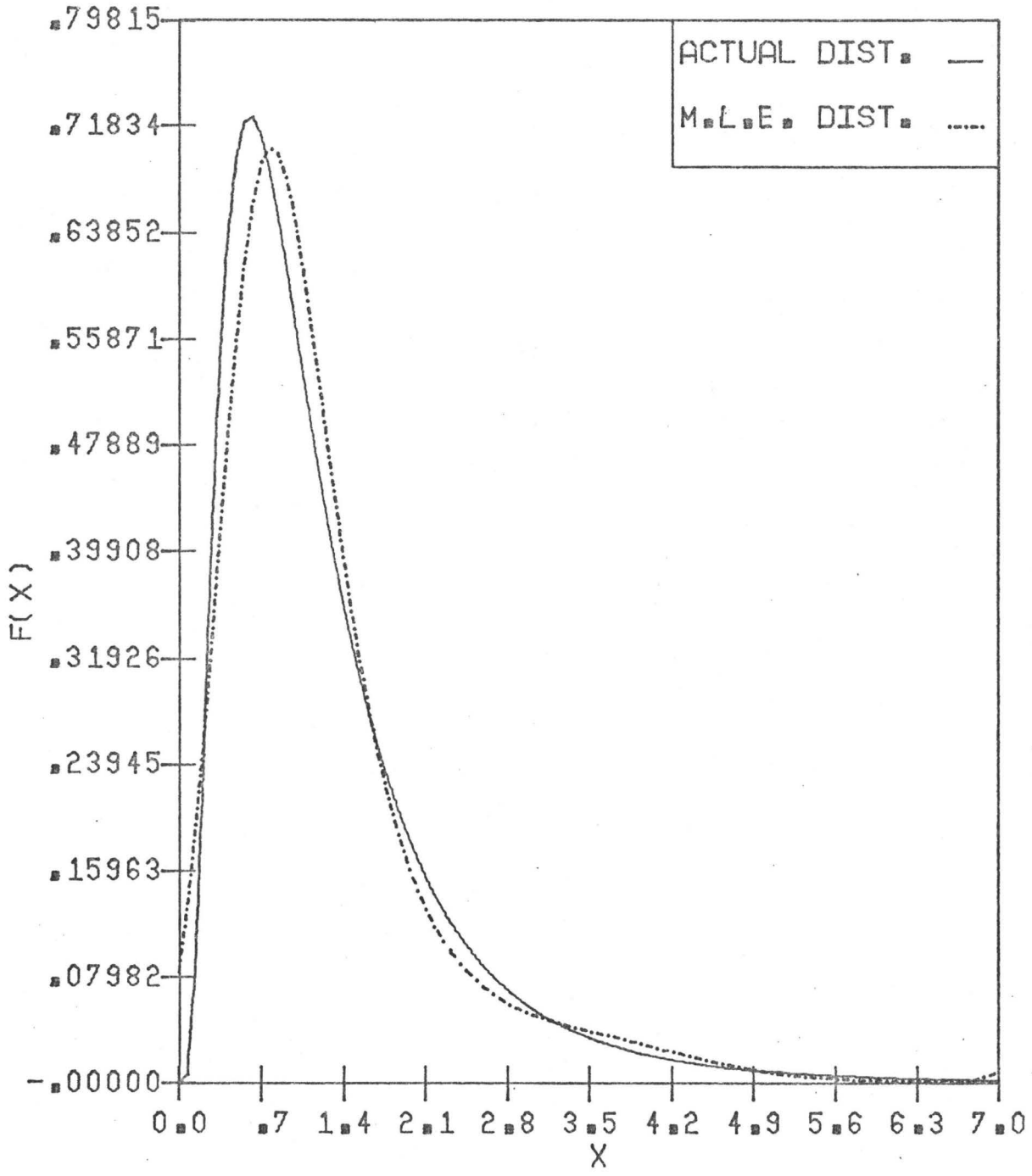


Figure A.32 Approximating Log-Normal distribution (LN3) ($\mu = 0.0$, $\sigma^2 = 1.0$), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: 1.261620 .888137 1.63326 6.455104 24.81569

λ values (for M.L.E. Distribution): - 2.4648 6.4458 - 6.36595 + 2.284981 - .36131106 .020727

Percentage area in common between the two distributions = 92.84

LOG-NORMAL DISTRIBUTION (LN4)

$$f(x) = \begin{cases} \frac{1}{\sigma \Pi \ 2\Pi} \exp\left[-\frac{1}{2\Pi^2}(\log x - \mu)^2\right], \ \mu = 0, \ \sigma^2 = .3, \ x \geq 0 \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 1.0773 .18610 .09725

.18845 .27287

Standardized moment measures ($\sqrt{\beta_1}$, β_2): 1.211 5.441

Upper and lower bounds: 4.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.32 Common area between Log-Normal Distribution (LN4), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments.	Percentage of area in common between the two curves
3	90.91
4	96.02
5	97.70

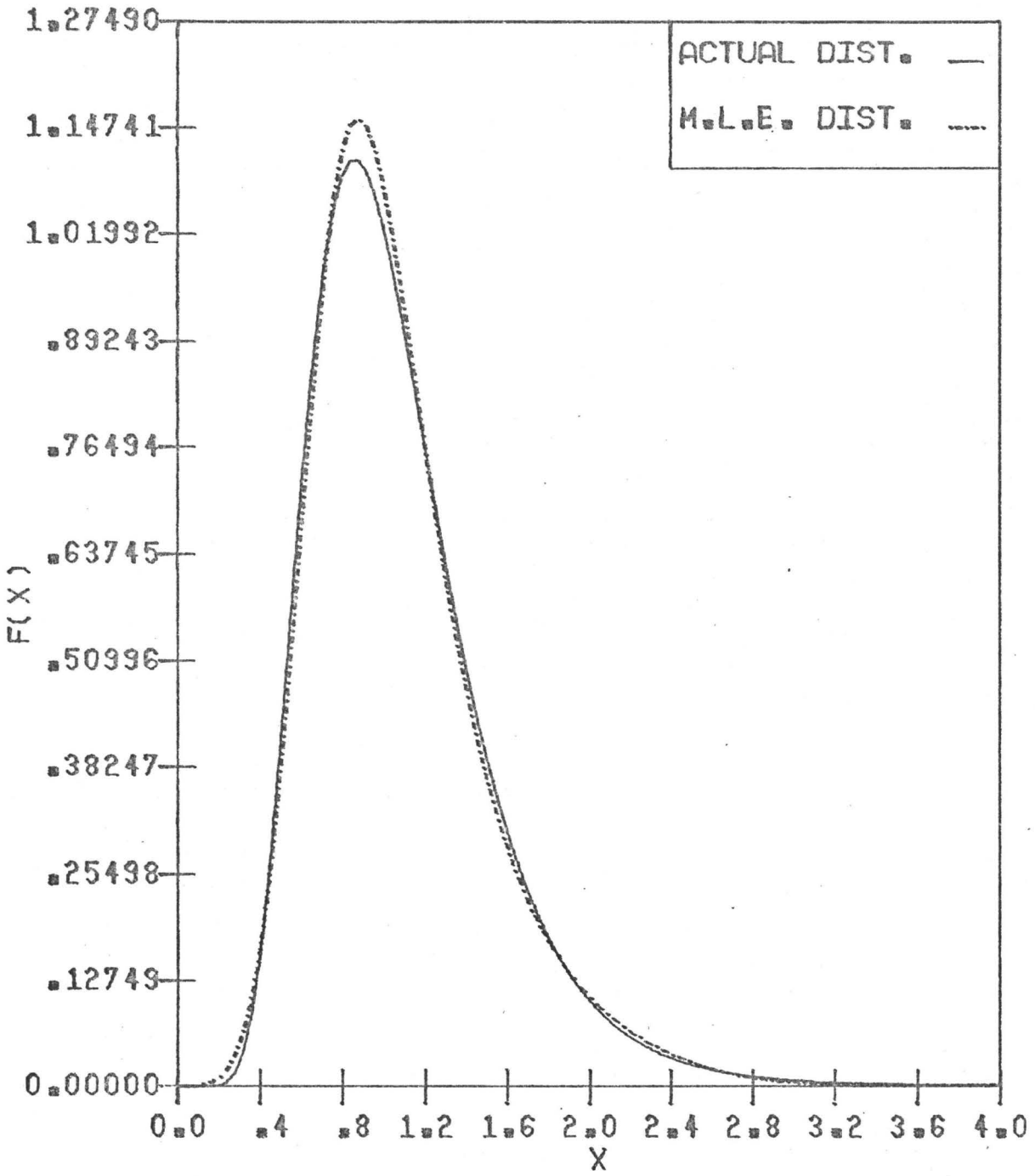


Figure A.33 Approximating Log-Normal Distribution (LN4) ($\mu = 0$, $\sigma^2 = .3$), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first five moments.

Central Moment values: 1.07730 .18610 .097237 .18845 .272869

λ values (for M.L.E. Distribution): - 9.4254 29.9921 - 33.7175 16.9511 - 4.10425 .37979

Percentage area in common between the two distributions = 97.70

LOG-NORMAL DISTRIBUTION (LN5)

$$f(x) = \begin{cases} \frac{1}{\sigma \Pi \ 2\Pi} \exp\left[-\frac{1}{2\Pi^2}(\log x - \mu)^2\right], \ \mu = 0, \ \sigma^2 = .1, \ x > 0 \\ \\ 0, \text{ elsewhere} \end{cases}$$

Central moment values: 1.02531 .05390 .00864 .011207

Standardized moment measures ($\sqrt{\beta_1}$, β_2): .631 3.245

Upper and lower bounds: 3.00 0.00

Type of curve: $\cap$ (bell shaped)

Table A.33 Common area between Log-Normal Distribution (LN5), and its approximate Maximum-Logarithmic Entropy Distribution for different known first moments.

Number of known first moments	Percentage of area in common between the two curves
3	95.64
4	98.87

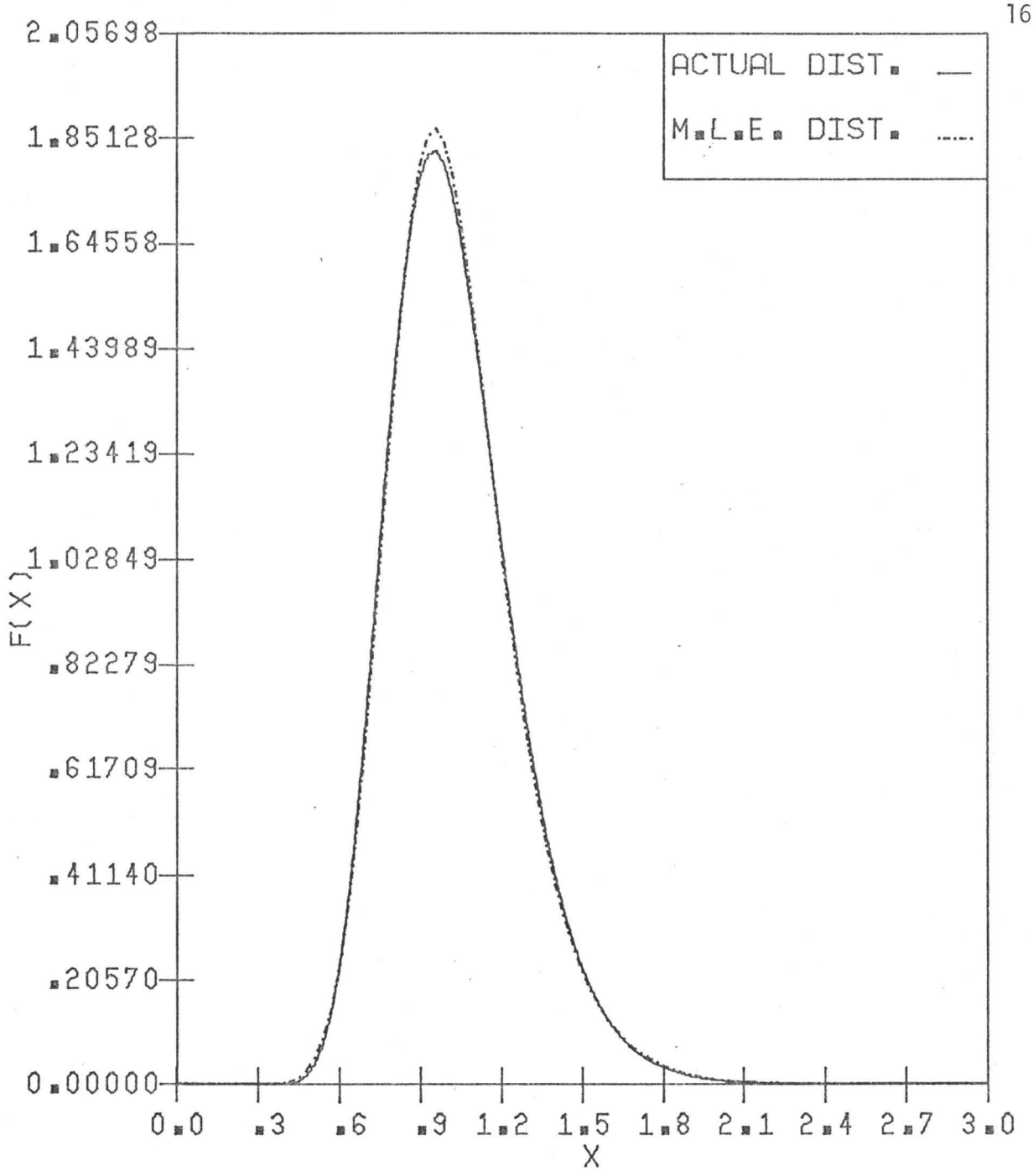


Figure A.34 Approximating Log-Normal Distribution (LN5) ($\mu = 0$, $\sigma^2 = .1$), by a Maximum-Logarithmic Entropy Distribution (M.L.E.), based on the first four moments.

Central Moment values: 1.02531 .053898 .008642 .011207

λ values (for M.L.E. Distribution): - 24.8617 72.5883 - 72.5529 30.2225 - 4.79349

Percentage area in common between the two distributions = 98.87

APPENDIX B

DERIVATION OF $S = -\Sigma\, p_i \log p_i$ (10)

Let $S(\frac{1}{n}, \frac{1}{n}, \ldots \frac{1}{n}) = A(n)$. From condition (3) we can decompose a choice from S^m equally likely possibilities into a series of m choices each from S equally likely possibilities and obtain

$$A(S^m) = m A(S)$$

Similarly

$$A(t^n) = n A(t)$$

We can choose n arbitrarily large and find an m to satisfy

$$S^m \leqslant t^n < S^{(m+1)}$$

Taking logarithms and dividing by n log S, gives

$$\frac{m}{n} \leqslant \frac{\log t}{\log S} \leqslant \frac{m}{n} + \frac{1}{n} \qquad \text{or}$$

$$\left| \frac{m}{n} - \frac{\log t}{\log S} \right| < \varepsilon$$

where ε is arbitrarily small. Now from the monotonic property of A (n)

$$A(S^m) \leqslant A(t^n) \leqslant A(S^{m+1})$$

$$m A(S) \leqslant n A(t) \leqslant (m+1) A(S)$$

Dividing by n A (S) gives

$$\frac{m}{n} \leqslant \frac{A(t)}{A(S)} \leqslant \frac{m}{n} + \frac{1}{n} \qquad \text{or}$$

$$\left| \frac{m}{n} - \frac{A(t)}{A(S)} \right| < \varepsilon$$

$$\left| \frac{A(t)}{A(S)} - \frac{\log t}{\log S} \right| \leqslant 2\varepsilon$$

$$A(t) = -K \log t$$

where K must be positive to satisfy condition (2). Now suppose we have a choice from n possibilities with commeasurable probabilities

$$p_i = \frac{n_i}{\Sigma n_i}$$

where the n_i are integers. We can break down a choice from Σn_i possibilities into a choice from n possibilities with probabilities $p_i, \ldots . p_n$ and then, if the ith is chosen, a choice from n_i with equal probabilities. Using condition (3) again, we equate the total choice from Σn_i as computed by two methods

$$K \log \Sigma n_i = H(p_1, \ldots , p_n) + K \Sigma p_i \log n_i$$

Hence

$$H = K\,[\Sigma\, p_i \log \Sigma n_i - \Sigma\, p_i \log n_i]$$

$$= K\, p_i \log \frac{n_i}{\Sigma n_i} = - K\, \Sigma\, p_i \log p_i$$

If the p_i are incommeasurable, they may be approximated by rationals and the same expression holds in general. The choice of coefficient K is a matter of convenience and amounts to the choice of a unit of measure.

APPENDIX C

COMPUTER PROGRAMS AND USER'S MANUAL

HOW TO USE

1. Write calling program. In its simplest form it is as follows.

(a) DIMENSION statement. Check through the list of input, and output variables. Include all subscripted variables, dimensioning as indicated.

(b) Define input data. Include DATA cards, or READ statements, or individual cards such as

```
IPRINT = 1
```

so that each variable in the input list is defined.

(c) Call the subroutine. For example

```
CALL DECI1(CM,XMIN,XMAX,X,XP,NP,NXP,IPRINT,IRESULT,N)
```

(d) Add STOP and END

2. If DECI1 or MOMENT are called, the user must write subroutine DERV to evaluate $g(x_1, x_2, \ldots, x_n)$,

$$\frac{\partial g}{\partial x_i} \quad \text{and} \quad \frac{\partial^2 g}{\partial x_i^2}.$$

3. Add to the deck all subroutines called as indicated in the documentation.

CONTENTS

SUBROUTINE DECI1(CM,XMIN,XMAX,X,XP,YP,NXP,
IPRINT,IRESULT,N)

Purpose

Analog prediction under risk.

This subroutine provides an estimate of the probability density function for y, where

$$y = g(x_1, x_2, \ldots\ldots, x_n)$$

and the first four moments of all the x's are known.

The density function of y has the form

$$f(y) = \exp(\lambda_1 + \lambda_2 y + \lambda_3 y^2 + \lambda_4 y^3 + \lambda_5 y^4)$$

The program gives the values of the λ's. It also provides values of the cumulative distribution function for given values of the independent variable y.

Method

The first four moments of y are approximated in terms of the moments of the x_i's by using a truncated Taylor's series expansion[1]. Then the method of maximum entropy[2] is applied to generate the distribution. The relationship between the subroutines is illustrated in Figure 1. The user calls DECI1, which calculates the first four moments, and the lower and the upper bounds of the function

$$f(y) = g(x_1, x_2, \ldots\ldots, x_n),$$

using subroutine DERV, which is supplied by the user. Then DECI1 calls internally subroutine MEP to generate the distribution. In Figure 1 the user supplied routines are shown in double lines.

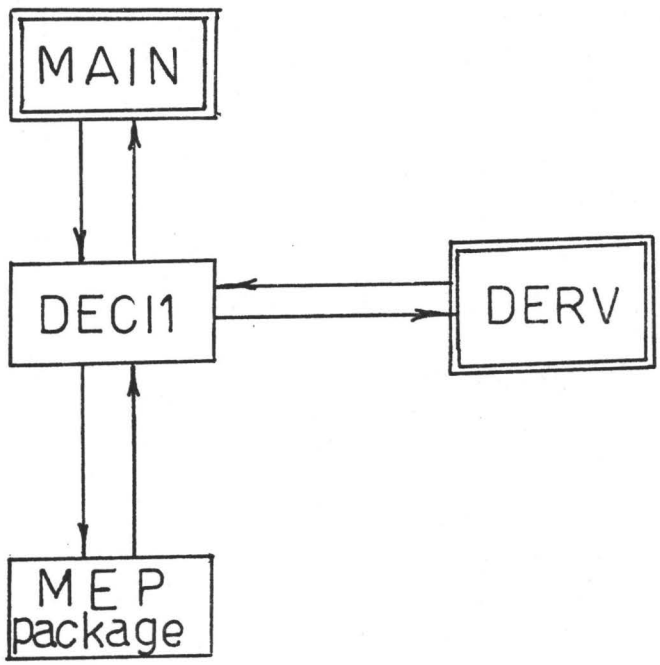


FIG. 1: The relationship between the subroutines. The user supplied routines are shown in double lines.

References

1. Siddall, J.N.; Analytical Decision Making in Engineering Design, Prentice-Hall, 1972.

2. Diab, Y.; The Maximization of the Logarithmic Entropy Function as a New Effective Tool in Statistical Modeling and Analytical Decision Making, Masters Thesis 1972, McMaster University.

Input Variables

N	number of independent variables.
CM(I,J)	array containing the first four moments of the independent variables, dimensioned (N,4).
XMIN(I)	lower bounds of the independent variables, dimensioned with the value of N.
XMAX(I)	upper bounds of the independent variables, dimensioned with the value of N.
NXP	number of points for which it is required to calculate the cumulative distribution function.
XP(I)	array containing the values of y for which the values of the cumulative distribution function are to be calculated, dimensioned with the value of NXP.
IPRINT	prints results every IPRINT cycle, set = 0 for no intermediate output. (Note--the intermediate results are related to the entropy maximization method, and have no direct relevance to analog prediction under risk. See subroutine MEP).

IDATA = 1, all input data is printed out.

= 0, input data is not printed out.

IRESULT = 1, output data is printed and plotted.

= 0, no output.

Output Variables

X(I) array containing the Lagrangian multipliers or λ's, dimension at 6. (Note--although there are five λ's the sixth subscript is used internally).

YP(I) array containing the values of the cumulative distribution function of f(y) corresponding to XP(I), dimensioned with the value of NXP.

Programming Information

DECI1 has full variable dimensioning. The calling program must provide dimensioning as given above.

The user must define the function $g(x_1,x_2,....,x_n)$, and the first and second partial derivatives. See SUBROUTINE DERV below.

SUBROUTINE DERV(FUN,DE1,DE2,N,X)

Purpose

To evaluate $g(x_1,x_2,....,x_n)$, $\frac{\partial g}{\partial x_i}$ and $\frac{\partial^2 g}{\partial x_i^2}$.

Method

These may be evaluated in any manner, including numerical approximations, as long as specific values are returned for any input point for the x's.

Input Variables

X(I)	point at which functions are to be evaluated.
N	number of x's.

Output Variables

FUN	function value or value of $g(x_1,x_2,\ldots,x_n)$.
DE1(I)	array giving value of first derivatives, $\frac{\partial g}{\partial x_i}$.
DE2(I)	array giving value of second derivatives, $\frac{\partial^2 g}{\partial x_i^2}$

How to Set Up Subroutine DERV

The following cards must be punched by the user:

```
SUBROUTINE DERV(FUN,DE1,DE2,N,X)
DIMENSION X(1),DE1(1),DE2(1)

Coding to define FUN, DE1(I), and DE2(I).
It may include any legal FORTRAN statements
and call to auxiliary subroutines.
RETURN
END
```

Listing

The following listing is for subroutine DECI1; for subroutine MEP (which is called by DECI1, to generate PDF) and its auxiliary subroutines; see MEP user's manual below.

```
      SUBROUTINE DECI1 (CM,XMIN,XMAX,X,XP,YP,NXP,IPRINT,IDATA,IRESULT,N)
      DIMENSION CM(N,4), XMIN(1), XMAX(1), DE1(4), DE2(4), CC(4), XP(1),
     1 YP(1),X(1)
      IF (IDATA.EQ.0) GO TO 2
      WRITE (6,11)
      WRITE (6,12) IDATA
      WRITE (6,13) IPRINT
      WRITE (6,14) IRESULT
      WRITE (6,15) N
      WRITE (6,16)
      DO 1 I=1,N
      WRITE (6,17) I,(CM(I,J),J=1,4),XMIN(I),XMAX(I)
1     CONTINUE
2     CONTINUE
      IDTA=1
      IF (IPRINT.EQ.0) IDTA=0
      DO 3 I=1,N
      CC(I)=1.
3     CONTINUE
      CALL DERV (FUN,DE1,DE2,N,CC)
      DO 4 I=1,N
      CC(I)=XMIN(I)
      IF (DE1(I).GT.0.0) CC(I)=XMAX(I)
4     CONTINUE
      CALL DERV (XTMAX,DE1,DE2,N,CC)
      DO 5 I=1,N
      CC(I)=XMAX(I)
      IF (DE1(I).GT.0.0) CC(I)=XMIN(I)
5     CONTINUE
      CALL DERV (XTMIN,DE1,DE2,N,CC)
      DO 6 I=1,N
      CC(1)=CM(I,1)
6     CONTINUE
      CALL DERV (FUN,DE1,DE2,N,CC)
      CC(1)=FUN
      DO 7 I=2,4
      CC(I)=0.0
7     CONTINUE
      DO 10 I=1,N
      CC(1)=CC(1)+.5*(DE2(I)*CM(I,2))
      CC(2)=CC(2)+DE1(I)**2*CM(I,2)+DE1(I)*DE2(I)*CM(I,3)
      CC(3)=CC(3)+DE1(I)**3*CM(I,3)
      SUM=0.0
      KJ=I+1
      IF (KJ.GT.N) GO TO 9
      DO 8 J=KJ,N
      SUM=SUM+6.*(DE1(I)*DE1(J))**2*CM(I,2)*CM(J,2)
8     CONTINUE
9     CONTINUE
      CC(4)=CC(4)+SUM+DE1(I)**4*CM(I,4)
```

```
10    CONTINUE
      KSTART=1
      TOL=1.E-6
      CALL MEP (XTMAX,XTMIN,CC,4,X,XP,YP,NXP,IDTA,IPRINT,IRESULT,KSTART,
     1TOL)
      RETURN
C
11    FORMAT (1H1,//,20X,*INPUT DATA FOR SUBROUTINE   DECI1*,/,20X,31(*-
     1*),//)
12    FORMAT (* INPUT DATA IS PRINTED OUT FOR IDATA =1 ONLY . . .IDATA =
     1*,I18,/)
13    FORMAT (* INTERMEDIATE OUTPUT EVERY IPRINT(TH) CYCLE  . . IPRINT =
     1*,I18,/)
14    FORMAT (* OUTPUT DATA IS PRINTED OUT FOR IRESULT =1 ONLY IRESULT =
     1*,I18,/)
15    FORMAT (* NUMBER OF INDEPENDANT VARIABLES . . . . . N=
     1*,I18,/)
16    FORMAT (//,* VARIABLE    FIRST MOMENT      SECOND MOMENT      THIRD
     1 MOMENT      FOURTH MOMENT          LOWER LIMIT       HIGHER LIMIT
     2*,//)
17    FORMAT (1X,I3,6X,4E18.9,5X,2E18.9,/)
      END
```

SUBROUTINE MEP(XMAX,XMIN,CC,N,AL,XP,YP,NXP,
IDATA,IPRINT,IRESULT,KSTART,TOL)

Purpose

This subroutine provides an estimate of the probability density function for the random variable x, where the first n moments of x are known. The density function, y, has the form

$$y = \exp(\lambda_1 + \lambda_2 x + \lambda_3 x^2 + ... + \lambda_{n+1} x^n) \qquad .....(1)$$

The program gives the values of the λ's. It also provides values of the cumulative distribution function for a given value of the independent variable x.

Method

Langrange's method of undetermined multipliers is used to maximize Shannon's Logarithmic Entropy Function ($p_i \ell n p_i$), with the given n moments as constraints. This leads to n algebraic simultaneous equations in n unknowns, where the unknowns are the λ_i's ($i \neq 1$) in equation (1), and only one solution exists. The n equations are solved by optimizing (minimizing) the square of the relative error in the value of the moments; the optimization process stops when the error becomes less than the accuracy specified by the user (TOL). The Jacobson-Oksman algorithm is used. The program provides internally a starting point to start the algorithm. Four methods are used; if one fails to provide a solution, the next is called automatically without participation of the user.

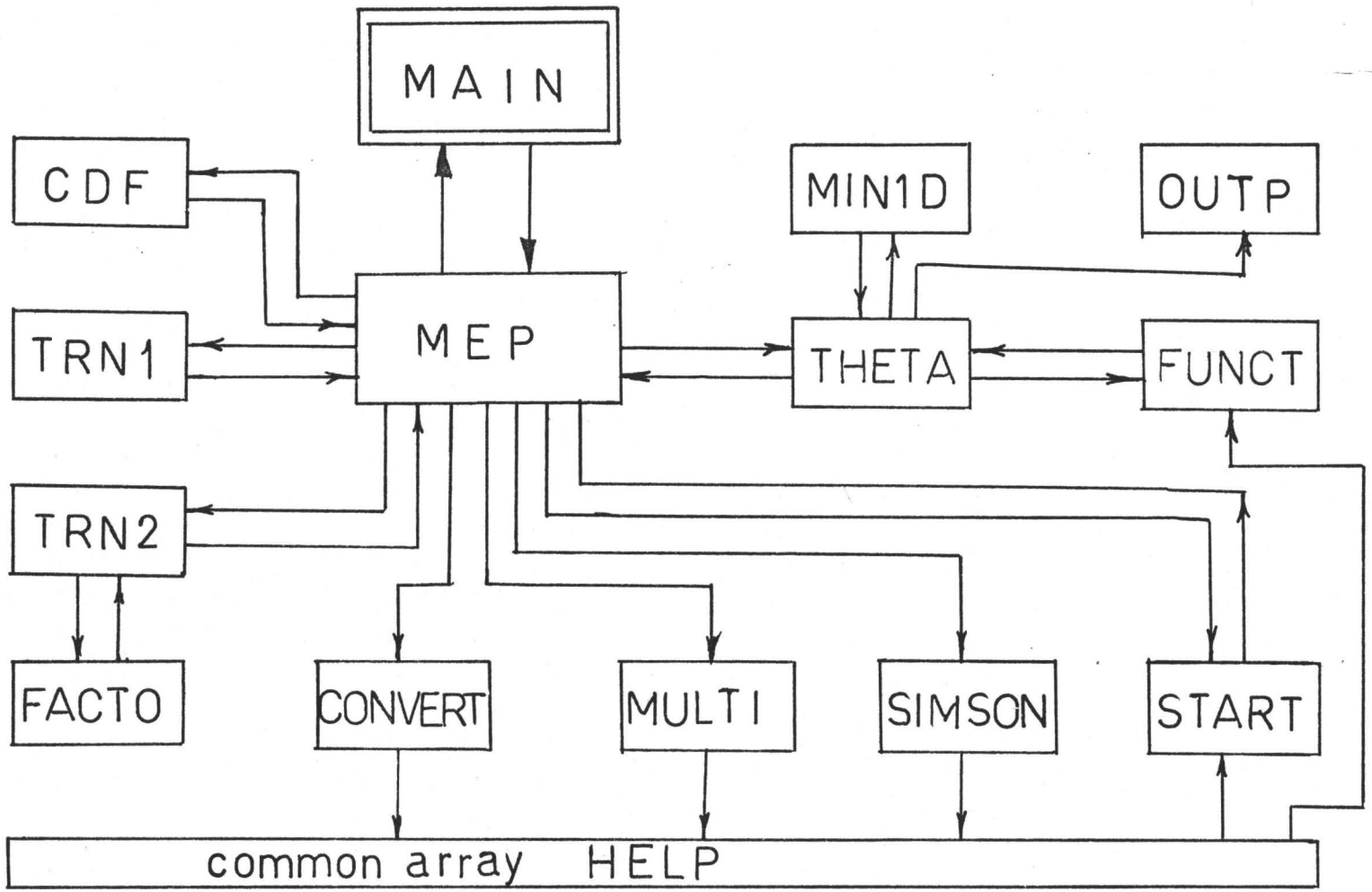


FIG. 1: Relation between Subroutine MEP, MAIN and the auxiliary subroutines. The user supplied routine is shown in double lines.

To eliminate overflow, the problem is converted from its original domain, (XMAX - XMIN), to a modified domain,(XMAX = 1, XMIN = 0). When a solution is found, the problem is transferred again to its original domain. SUBROUTINES TRN1, TRN2 and CONVERT look after this transfer problem.

References

1. Diab, Y.; "The Maximization of the Logarithmic Entropy Function as a New Effective Tool in Statistical Modeling and Analytical Decision Making", Masters Thesis 1972, McMaster University.

2. Jacobson, D.H. and Oksman, W.; "An Algorithm that Minimizes Homogeneous Functions of N variables in N+2 Iterations and Rapidly Minimizes General Functions", Technical Report No. 618, Oct. 1970, Div. of Engineering and Applied Physics, Harvard University, Cambridge, Mass.

Input Variables

N — number of first moments, should be less than or equal to 6.

CC(I) — array containing the first N moments, dimensioned (N).

XMIN — lower bound of the variable.

XMAX — upper bound of the variable.

TOL — the allowed relative error in the moment value, a solution exists where

$$R(I) = \frac{cc_i - cc_i'}{c_i} \leq TOL \qquad .....(2)$$

where cc_i is the given moment

cc_i' is the predicted moment.

A reasonable value for TOL is 10^{-6}.

IDATA = 1, all input data is printed out

= 0, input data is not printed out.

IPRINT prints results every IPRINT cycle, set = 0 for no intermediate output. If IPRINT $\neq$ 0, all intermediate results before optimization, the starting method name, and the starting values of the λ's, are printed out. In addition the following are printed, cycle number, number of function evaluations (subroutine FUNCT), the normgradient, total residuals ($\sum_{i}^{n} R_i^2$) where R_i is defined in equation (1), the values of the λ's, and the value of each individual R_i.

IRESULT = 1, output data is printed and plotted.

= 0, no output.

KSTART = 1, normal assumption starting method.

= 2, uniform assumption starting method

= 3, N points assumption starting method.

= 4, step by step assumption starting method.

Set to 1, in case no particular starting method is preferred. In this event the subroutine will try other methods if one method fails.

NXP number of points for which it is required to calculate the cumulative distribution function.

XP array containing the values of the independent variable for which the values of the cumulative distribution functions are to be calculated, dimensioned with the value of NXP.

Output Variables

AL(I) array containing the Lagrangian multipliers or λ's, dimensioned at (N+2) (Note--although there are only (N+1) λ's, the (N+2) (th) subscript is used internally).

YP(I) array containing the values of the cumulative distribution function of XP, dimensioned with the value of NXP.

Programming Information

MEP has full variable dimensioning. The calling program must provide dimensioning as given above.

Listing

The following listing is for subroutine MEP and the auxiliary subroutines.

```
      SUBROUTINE MEP (XMAX,XMIN,CC,N,AL,XP,YP,NXP,IDATA,IPRINT,IRESULT,K
     1START,TOL)
C
      COMMON /FAIL/ NFAIL
      COMMON /HELP/ S(31),XX(8,31),C(8),M
      DIMENSION AL(1), CC(1), ETA(4), XP(1), YP(1)
C
C     WRITE THE INPUT DATA
C
      IF (IDATA.EQ.0) GO TO 1
      WRITE (6,22)
      WRITE (6,23) IDATA
      WRITE (6,24) IPRINT
      WRITE (6,25) IRESULT
      WRITE (6,26) N
      WRITE (6,27) XMAX
      WRITE (6,28) XMIN
      WRITE (6,29) (CC(I),I=1,4)
      IF (N.GT.4) WRITE (6,19) (CC(I),I=5,N)
      WRITE (6,30) TOL
      WRITE (6,31) NXP
1     CONTINUE
      NFAIL=0
      AL(N+1)=2.
      AL(N+2)=0.0
      M=31
      X2MIN=0.0
      X2MAX=1.
C
C     CALCULATE THE MOMENTS AT THE MODIFIED LIMITS
C
      CALL TRN1 (XMAX,XMIN,CC,X2MAX,X2MIN,N)
C
C     CALCULATE THE MOMENTS ABOUT THE ORIGIN FOR THE MODIFIED LIMITS ,
C     STORE THEM IN  HELP  COMMON ARRAY
C
      CALL CONVERT (CC,N)
C
C     GENERATE THE SIMPSON MULTIPLIERS AND STORE THEM IN HELP COMMON
C
      CALL SIMSON
C
C     GENERATE THE X,S POWER FOR SUBROUTINE FUNCT, STORE THEM IN HELP
C     COMMON ARRAY
C
      CALL MULTI (X2MAX,X2MIN,N)
C
C     DEFINE THE INPUT DATA FOR SUBROUTINE THETA
C
      ETA(1)=1.E-12
```

```
      ETA(2)=TOL
      ETA(3)=1.E-24
      ETA(4)=1.E-24
      MAXFN=1000
      MODE=1
      UMIN=0.0
C
C     WRITE THE INTERMEDIATE RESULTS YOU HAVE OBTAINED SO FAR
C
      IF (IPRINT.EQ.0) GO TO 2
      WRITE (6,32)
      WRITE (6,33) M
      WRITE (6,34) X2MAX,X2MIN
      WRITE (6,35) (CC(I),I=1,4)
      IF (N.GT.4) WRITE (6,20) (CC(I),I=5,N)
      WRITE (6,36) (C(I),I=1,4)
      IF (N.GT.4) WRITE (6,20) (C(I),I=5,N)
      WRITE (6,37) (ETA(I),I=1,4)
2     CONTINUE
C
C     FIND A STARTING POINT FOR SUBROUTINE THETA TO START THE OPTIMIZAT-
C     ION ALGORITHM
C
      IF (KSTART.EQ.4) WRITE (6,42)
      CALL START (X2MAX,X2MIN,AL,KSTART,CC,N,IPRINT,UMIN,MODE,MAXFN,ETA)
      IF (NFAIL.EQ.1) GO TO 9
C
C     PRINT THE STARTING VALUES
C
      IF (IPRINT.EQ.0) GO TO 7
      GO TO (3,4,5,6), KSTART
3     WRITE (6,38)
      WRITE (6,39) (AL(I),I=1,4)
      IF (N.GT.4) WRITE (6,20) (AL(I),I=5,N)
      GO TO 7
4     WRITE (6,40)
      WRITE (6,39) (AL(I),I=1,4)
      IF (N.GT.4) WRITE (6,20) (AL(I),I=5,N)
      GO TO 7
5     WRITE (6,41)
      WRITE (6,39) (AL(I),I=1,4)
      IF (N.GT.4) WRITE (6,20) (AL(I),I=5,N)
      GO TO 7
6     WRITE (6,42)
      WRITE (6,39) (AL(I),I=1,4)
      IF (N.GT.4) WRITE (6,20) (AL(I),I=5,N)
7     CONTINUE
      NFAIL=0
      IF (IPRINT.EQ.0) GO TO 8
      WRITE (6,43)
```

```
8       CONTINUE
        CALL THETA (AL,N,ETA,UMIN,MAXFN,MODE,IPRINT)
        IF (NFAIL.EQ.0) GO TO 10
        IF (KSTART.EQ.4) GO TO 9
C
C       THE PROGRAM HAS FAILED SO FAR , TRY ANOTHER STARTING POINT AND TRY
C       AGAIN
C
        KSTART=KSTART+1
        IF (KSTART.EQ.4.AND.N.LE.2) GO TO 9
        GO TO 2
9       CONTINUE
        WRITE (6,44)
        CALL EXIT
10      CONTINUE
C
C       CALCULATE THE ZEROTH LAGRANGIAN MULTIPLIER
C
        SUM=0.0
        DO 12 I=1,M
        SZ=0.0
        DO 11 K=1,N
        SZ=SZ+AL(K)*XX(K,I)
11      CONTINUE
        SUM=SUM+S(I)*EXP(SZ)
12      CONTINUE
        NPL=N+1
        DO 13 I=1,N
        K=N+2-I
        AL(K)=AL(K-1)
13      CONTINUE
        DELTA=(X2MAX-X2MIN)/FLOAT(M-1)
        AL(1)=-ALOG(SUM*DELTA/3.)
        IF (IPRINT.EQ.0) GO TO 14
        WRITE (6,45) (AL(I),I=1,NPL)
14      CONTINUE
C
C       CALCULATE THE LAGRANGIAN MULTIPLIERS FOR THE ORIGINAL LIMITS
C
        CALL TRN2 (XMAX,XMIN,AL,X2MAX,X2MIN,N)
C
C       CALCULATE THE CUMULATIVE DISTRIBUTION FUNCTION VALUE AT THE GIVEN
C       POINT
C
        DO 15 I=1,NXP
        YP(I)=CDF(XMIN,XMAX,XP(I),AL,N)
15      CONTINUE
        IF (IRESULT.NE.1) RETURN
C
C       PRINT AND PLOT THE RESULT
```

```
C
      WRITE (6,46)
      WRITE (6,47)
      WRITE (6,48) (AL(I),I=1,NPL)
      WRITE (6,21)
      DO 16 I=1,NXP
      WRITE (6,18) I,XP(I),YP(I)
16    CONTINUE
C
C     PLOT THE DISTRIBUTION
C
      M=51
      DELTA=(XMAX-XMIN)/FLOAT(M-1)
      DO 17 I=1,M
      X=XMIN+FLOAT(I-1)*DELTA
      Y=ENTRPF(AL,N,X)
      CALL PLOTPT (X,Y,9)
17    CONTINUE
      CALL OUTPLT
      RETURN
C
C
C
18    FORMAT (I10,12X,E18.9,12X,E18.9)
19    FORMAT (57X,4E18.9,//)
20    FORMAT (57X,4E18.9,//)
21    FORMAT (///,*          N                              VARIABLE
     1     CUMULATIVE  */58X,*DISTRIBUTION*,//)
22    FORMAT (1H1,//,20X,*INPUT DATA FOR SUBROUTINE   MEP*,/,20X,31(*-*)
     1,//)
23    FORMAT (* INPUT DATA IS PRINTED OUT FOR IDATA =1 ONLY . . .IDATA =
     1*,I18,/)
24    FORMAT (* INTERMEDIATE OUTPUT EVERY IPRINT(TH) CYCLE  . . IPRINT =
     1*,I18,/)
25    FORMAT (* OUTPUT DATA IS PRINTED OUT FOR IRESULT =1 ONLY IRESULT =
     1*,I18,/)
26    FORMAT (* NUMBER OF KNOWN FIRST MOMENTS . . . . . N=
     1*,I18,/)
27    FORMAT (* HIGHER LIMIT  . . . . . XMAX =
     1*,E18.9,/)
28    FORMAT (* LOWER LIMIT . . . . . XMIN =
     1*,E18.9,/)
29    FORMAT (* FIRST MOMENTS . . . . . CC(I) =
     1*,4E18.9,/)
30    FORMAT (* THE ALLOWED TOLERANCE IN LAGRANGIAN EQUATIONS . . .TOL =
     1*,E18.9,/)
31    FORMAT (* THE CUMULATIVE DISTRIBUTION REQUIRED AT NXP POINTS.NXP =
     1*,I18,/)
32    FORMAT (1H1,//,20X,*INTERMEDIATE RESULTS FOR SUBROUTINE   MEP*,/,2
     10X,41(*-*),//)
```

```
33    FORMAT (* NUMBER OF INTEGRATION STATION . . . . . M =
     1*,I18,/)
34    FORMAT (* MODIFIED MAXIMUM AND MINIMUM LIMITS . .  X2MAX , X2MIN =
     1*,2E18.9,/)
35    FORMAT (* MODIFIED MOMENTS ABOUT THE EXPECTED VALUE . . . .CC(I) =
     1*,4E18.9,/)
36    FORMAT (* MODIFIED MOMENTS ABOUT THE ORIGIN . . . . . C(I) =
     1*,4E18.9,/)
37    FORMAT (* SUBROUTINE THETA TOLERANCES . . . . . ETA(I) =
     1*,4E18.9,/)
38    FORMAT (//,* NORMAL ASSUMPTION STARTING METHOD*/34(*-*),/)
39    FORMAT (* STARTING VALUES   . . . . . AL(I) =
     1*,4E18.9,/)
40    FORMAT (//,* UNIFORM ASSUMPTION STARTING METHOD*/35(*-*),/)
41    FORMAT (//,* N POINTS STARTING METHOD*/25(*-*),/)
42    FORMAT (//,* STEP BY STEP STARTING METHOD*/29(*-*),/)
43    FORMAT (//,*  CYC NUMF   NORMGRAD          TOTAL*,24X,*VARIABLES*,40
     1X,*RESIDUALS*,/,*  NO.*,22X,*RESIDUALS        X(1)            X(2)
     2        X(3)            X(4)        R(1)        R(2)        R(3)        R
     3(4)*,//)
44    FORMAT (* THE PROGRAM HAS FAILED*)
45    FORMAT (* THE MODIFIED LAGRANGIAN MULTIPLIERS ARE . . . . . .
     1*,4E18.9/57X,4E18.9)
46    FORMAT (1H1,20X,*RESULTS FOR SUBROUTINE  DECI1*,/,20X,29(*-*),//)
47    FORMAT (* THE MATHEMATICAL MODEL OF THE MAXIMUN ENTROPY PROBABILIT
     1Y DISTRIBUTION AS THE FORM*,///,10X,60H Y=EXP(Z(1)+Z(2)*X+....+...
     2.+Z(I+1)*X**I+...+..+Z(N+1)*X**N)////,* WHERE          X          I
     3S THE VARIABLE*//,15X,*Y          IS THE CORRESPONDING PROBABILITY
     4 DENSITY FUNCTION*//,15X,*Z(I)       ARE CONSTANTS EQUALS TO*,//)
48    FORMAT (/,25X,5E18.9,/,25X,5E18.9,/)
      END
```

```
      SUBROUTINE FUNCT (N,AL,U,GRAD,RR)
C
C     THIS  SUBROUTINE IS USED TO CALCULATE THE OPTIMIZATION AND THE
C     GRADIENT AT ANY GIVEN POINT FO SUBROUTINE THETA
C
      DIMENSION AL(1), GRAD(1), SUM(9), RR(1)
      COMMON /FAIL/ NFAIL
      COMMON /HELP/ S(31),XX(8,31),C(8),M
      N21=2*N+1
      ZERO=0.0
      DO 1 I=1,N21
      SUM(I)=0.0
1     CONTINUE
2     CONTINUE
      DO 4 I=1,M
      SZ=ZERO
      DO 3 K=1,N
      SZ=SZ+AL(K)*XX(K,I)
3     CONTINUE
      IF (SZ.GT.740.) GO TO 9
      SS=EXP(SZ)*S(I)
      SUM(1)=SUM(1)+SS
      DO 4 J=2,N21
      SUM(J)=SUM(J)+XX(J-1,I)*SS
4     CONTINUE
      DO 5 I=2,N21
      SUM(I)=SUM(I)/SUM(1)
5     CONTINUE
      U=0.0
      DO 6 I=1,N
      RR(I)=(SUM(I+1)-C(I))/C(I)
      U=U+RR(I)*RR(I)
6     CONTINUE
      DO 8 K=1,N
      GRAD(K)=0.0
      DO 7 J=1,N
      GRAD(K)=GRAD(K)+(SUM(J+K+1)-SUM(J+1)*SUM(K+1))*RR(J)/C(J)
7     CONTINUE
      GRAD(K)=GRAD(K)*2.
8     CONTINUE
      RETURN
9     CONTINUE
      AA=SZ-320.
      ZERO=ZERO-AA
      GO TO 2
C
      END
```

```
      SUBROUTINE START (XMAX,XMIN,ALAMDA,KSTART,CC,NL,IPRINT,UMIN,MODE,M
     1AXFN,ETA)
C
C     THIS SUBROUTINE IS USED TO FIND A REASONABLE STARTING POINT FOR
C     SUBROUTINE  THETA
C
      COMMON /HELP/ S(31),XX(8,31),C(8),M
      DIMENSION R(11)
      DIMENSION CC(1), ETA(1)
      DIMENSION ALAMDA(1), X(6), Y(6), W(6,6)
      COMMON /FAIL/ NFAIL
      GO TO (3,1,5,26), KSTART
 1    CONTINUE
      NFAIL=0
      DO 2 I=1,NL
      ALAMDA(I)=0.0
 2    CONTINUE
      RETURN
 3    CONTINUE
      NFAIL=0
      ALAMDA(1)=CC(1)/CC(2)
      ALAMDA(2)=-.5/CC(2)
      DO 4 I=3,NL
      ALAMDA(I)=0.0
 4    CONTINUE
      RETURN
 5    CONTINUE
      NFAIL=0
      NNN=NL/2
      NNN=NNN*2
      NP1=NL+1
      DELTA=(XMAX-XMIN)/FLOAT(NL)
      DO 6 I=1,NP1
      X(I)=XMIN+FLOAT(I-1)*DELTA
 6    CONTINUE
      IF (NNN.NE.NL) GO TO 19
      W(1,1)=W(1,NP1)=1.
      DO 7 I=2,NL,2
      W(1,I)=4.
 7    CONTINUE
      IF (NL.EQ.2) GO TO 9
      NM1=NL-1
      DO 8 I=3,NM1,2
      W(1,I)=2.
 8    CONTINUE
 9    CONTINUE
      DO 10 J=1,NP1
      DO 10 I=2,NP1
 10   W(I,J)=W(I-1,J)*X(J)
      Y(1)=3./DELTA
```

```
      DO 11 I=1,NL
      Y(I+1)=C(I)*Y(1)
11    CONTINUE
      CALL SOLVE (W,Y,ID,NP1,6)
12    CONTINUE
      DO 13 I=1,NP1
      DO 13 J=1,NP1
13    W(I,J)=.0
      DO 14 I=1,NP1
      IF (Y(I).LE.0.0) Y(I)=.0002
14    CONTINUE
      DO 15 I=1,NP1
      Y(I)=ALOG(Y(I))
15    CONTINUE
      DO 16 I=1,NP1
      W(I,1)=1.
16    CONTINUE
      DO 17 I=2,NP1
      DO 17 J=1,NP1
17    W(J,I)=W(J,I-1)*X(J)
      CALL SOLVE (W,Y,ID,NP1,6)
      DO 18 I=1,NL
      ALAMDA(I)=Y(I+1)
18    CONTINUE
      RETURN
19    CONTINUE
      R(1)=3./8.
      R(4)=3./8.
      R(2)=R(3)=9./8.
      IF (NL.EQ.3) GO TO 22
      R(NL+1)=1./3.
      R(4)=R(4)+1./3.
      DO 20 I=5,NL,2
      R(I)=4./3.
20    CONTINUE
      IF (NL.EQ.5) GO TO 22
      NS=NL-1
      DO 21 I=6,NS,2
      R(I)=2./3.
21    CONTINUE
22    CONTINUE
      DO 23 I=1,NP1
      W(1,I)=R(I)
23    CONTINUE
      DO 24 J=1,NP1
      DO 24 I=2,NP1
24    W(I,J)=W(1-1,J)*X(J)
      Y(1)=1./DELTA
      DO 25 I=1,NL
      Y(I+1)=C(I)*Y(1)
```

```
25    CONTINUE
      CALL SOLVE (W,Y,ID,NP1,6)
      GO TO 12
26    CONTINUE
      N=2
      ALAMDA(2)=-.5/CC(2)
      ALAMDA(1)=CC(1)/CC(2)
      NFAIL=0
27    CONTINUE
      ALAMDA(N+1)=2.0
      ALAMDA(N+2)=0.0
      CALL THETA (ALAMDA,N,ETA,UMIN,MAXFN,MODE,IPRINT)
      IF (NFAIL.EQ.1) RETURN
      IF (N.EQ.NL) RETURN
      ALAMDA(N+1)=0.0
      N=N+1
      GO TO 27
      END
```

```
      SUBROUTINE CONVERT (CM,NL)
C
C     THIS SUBROUTINE IS TO CALCULATE THE MOMENTS ABOUT THE ORIGIN
C
      COMMON /HELP/ S(31),XX(8,31),C(8),M
      DIMENSION CM(1)
      C(1)=CM(1)
      DO 2 J=2,NL
      C(J)=CM(J)-C(1)**J*(-1.)**J
      N=J-1
      DO 1 K=1,N
      C(J)=C(J)-(-1.)**K*FACTO(J)/(FACTO(K)*FACTO(J-K))*C(1)**(K)*C(J-K)
1     CONTINUE
2     CONTINUE
      RETURN
      END
```

```
      SUBROUTINE SIMSON
C
C     THIS SUBROUTINE IS TO CALCULATE THE SIMPSON MULTIPLIERS
C
      COMMON /HELP/ S(31),XX(8,31),C(8),M
      S(1)=1.
      S(M)=1.
      N=M-1
      DO 1 I=2,N,2
      S(I)=4.
1     CONTINUE
      N=N-1
      DO 2 I=3,N,2
      S(I)=2.
2     CONTINUE
      RETURN
      END
```

```
      SUBROUTINE MULTI (XMAX,XMIN,N)
C
C     THIS SUBROUTINE IS USED TO GENERATE THE X,S POWER FOR SUBROUTINE
C     FUNCT
C
      COMMON /HELP/ S(31),XX(8,31),C(8),M
      DELTA=(XMAX-XMIN)/FLOAT(M-1)
      DO 1 I=1,M
      XX(1,I)=XMIN+FLOAT(I-1)*DELTA
      NN=2*N
      DO 1 J=2,NN
      XX(J,I)=XX(J-1,I)*XX(1,I)
1     CONTINUE
      RETURN
      END
```

```
      SUBROUTINE TRN1 (X1MAX,X1MIN,C,X2MAX,X2MIN,NL)
C
C     THIS SUBROUTINE IS USED TO CALCULATE THE MOMENTS FOR THE MODIFIED
C     LIMITS
C
      DIMENSION C(1)
      SCL=(X1MAX-X1MIN)/(X2MAX-X2MIN)
      C(1)=C(1)/SCL-X1MIN/SCL+X2MIN
      DO 1 I=2,NL
      C(I)=C(I)/SCL**I
1     CONTINUE
      RETURN
      END
```

```
      SUBROUTINE TRN2 (X1MAX,X1MIN,X,X2MAX,X2MIN,N)
C
C     THIS SUBROUTINE IS USED TO CALCULATE THE LAGRANGIAN MULTIPLIERS
C     AT  THE ORIGINAL LIMITS
C
      DIMENSION X(1)
      S=(X1MAX-X1MIN)/(X2MAX-X2MIN)
      A=X2MIN-X1MIN/S
      X(1)=X(1)-ALOG(S)
      DO 1 I=1,N
      X(1)=X(1)+X(I+1)*A**I
1     CONTINUE
      DO 4 J=2,N
      DO 3 I=J,N
      FAC=1.
      KK=I-J+2
      DO 2 K=KK,I
      FAC=FAC*FLOAT(K)
2     CONTINUE
      X(J)=X(J)+FAC/FACTO(J-1)*A**(I-J+1)*X(I+1)
3     CONTINUE
      X(J)=X(J)/S**(J-1)
4     CONTINUE
      X(N+1)=X(N+1)/S**N
      RETURN
      END
```

```
      FUNCTION FACTO (M)
      FACTO=1.
      IF (M.EQ.0) RETURN
      DO 1 I=1,M
      FACTO=FACTO*FLOAT(I)
1     CONTINUE
      RETURN
      END
```

```
      FUNCTION FSIMP (FUNC,RANGE,M)
      DIMENSION FUNC(1)
      IF (M.EQ.1) GO TO 4
      XX=RANGE/(3.*FLOAT(M-1))
      AREA=FUNC(1)+FUNC(M)
      MM=M-1
      DO 1 I=2,MM,2
      AREA=AREA+4.*FUNC(I)
1     CONTINUE
      IF (M.EQ.3) GO TO 3
      MM=M-2
      DO 2 I=3,MM,2
      AREA=AREA+2.*FUNC(I)
2     CONTINUE
3     FSIMP=XX*AREA
      GO TO 5
4     FSIMP=0.0
5     RETURN
      END
```

```
      FUNCTION ENTRPF (AL,NL,X)
C
C     FUNCTION SUBROUTINE TO EVALUATE THE DISTRIBUTION ENTROPY FUNCTION
C     AT  A GIVEN POINT
C
      DIMENSION AL(1)
      NPL=NL+1
      S=AL(1)
      DO 1 I=2,NPL
      S=S+AL(I)*X**(I-1)
1     CONTINUE
      ENTRPF=EXP(S)
      RETURN
      END
```

```
      FUNCTION CDF (XMIN,XMAX,XP,AL,N)
C
C     THIS FUNCTION SUBROUTINE IS TO CALCULATE THE CUMMULATIVE DISTRIBU-
C     TION FUNCTION AT A GIVEN POINT
C
      DIMENSION AL(1)
      IF (XP.LE.XMIN) GO TO 3
      IF (XP.GE.XMAX) GO TO 4
      RANGE=XMAX-XMIN
      RANGEN=XP-XMIN
      SS=RANGEN/RANGE*51.
      JSS=SS
      JSS=(JSS/2)*2+5
      AREA=0.0
      JSM1=JSS-1
      DELTA=RANGEN/FLOAT(JSM1)
      DO 1 I=2,JSM1,2
      X=XMIN+FLOAT(I-1)*DELTA
      AREA=AREA+4.*ENTRPF(AL,N,X)
1     CONTINUE
      JSM1=JSM1-1
      DO 2 I=3,JSM1,2
      X=XMIN+FLOAT(I-1)*DELTA
      AREA=AREA+2.*ENTRPF(AL,N,X)
2     CONTINUE
      AREA=AREA+ENTRPF(AL,N,XMIN)+ENTRPF(AL,N,XP)
      AREA=AREA*DELTA/3.
      CDF=AREA
      GO TO 5
3     CDF=0.0
      GO TO 5
4     CDF=1.
5     CONTINUE
      RETURN
      END
```

```
      SUBROUTINE THETA (X,NDIM,ETA,EST,MAX,MODE,IPRINT)
      COMMON /FAIL/ NFAIL
      DIMENSION X(1), X1(6), X2(6), G1(6), G2(6), ALFA(6), H(6), P(6,6),
     1 Y(6), PY(6), PE(6), ETA(1), BIGV(6), RR(5)
      KTB=0
      IFLAG=0
      M=0
      N2=NDIM+1
      N1=NDIM+2
      NUMF=0
      IER=0
      DO 1 I=1,N1
      X1(I)=X(I)
1     CONTINUE
      CALL FUNCT (NDIM,X1,F1,G1,RR)
      NUMF=NUMF+1
      DO 2 I=1,NDIM
      X2(I)=X1(I)
      G2(I)=G1(I)
      H(I)=-G1(I)
2     CONTINUE
      F2=F1
      X2(N2)=X1(N2)
      X2(N1)=X1(N1)
3     CONTINUE
      KOUNT=0
      EPS=ETA(4)
      CALL MIN1D (FUNCT,X2,H,RO,NDIM,F2,G2,NUMF,IER,EPS,EST,RR,IPRINT)
      IF (NFAIL.EQ.1) RETURN
      IF (IER.NE.0) GO TO 30
      DO 4 I=1,N1
      BIGV(I)=X2(I)
      ALFA(I)=X2(I)
4     CONTINUE
      RO=-RO
      GG=0.
      DO 5 I=1,NDIM
      GG=GG+G2(I)*G2(I)
5     CONTINUE
      GG=SQRT(GG)
      IF (IPRINT.EQ.0) GO TO 7
      IF (MOD(KTB,IPRINT).NE.0) GO TO 6
      CALL OUTP (X2,F2,M,NDIM,GG,NUMF,RR)
6     KTB=KTB+1
7     DO 9 I=1,N1
      DO 8 J=1,N1
      P(I,J)=0.
8     CONTINUE
      P(I,I)=1.
9     CONTINUE
```

```
10    CONTINUE
      KOUNT=0
      KOUNT=KOUNT+1
11    DO 12 I=1,NDIM
      Y(I)=G2(I)
12    CONTINUE
      Y(N2)=F2
      Y(N1)=ETA(1)
      V=0.
      DO 13 I=1,NDIM
      V=V+X2(I)*G2(I)
13    CONTINUE
      YA=0.
      DO 14 I=1,N1
      YA=YA+Y(I)*ALFA(I)
14    CONTINUE
      VYA=V-YA
      BIGV(KOUNT)=V
      DO 15 I=1,N1
      PY(I)=0.
      PE(I)=P(I,KOUNT)
      DO 15 J=1,N1
15    PY(I)=PY(I)+P(J,I)*Y(J)
      EPY=PY(KOUNT)
      IF (ABS(EPY).LT.ETA(3)) GO TO 31
      PY(KOUNT)=PY(KOUNT)-1.
      DO 16 I=1,N1
      DO 16 J=1,N1
16    P(I,J)=P(I,J)-PE(I)*PY(J)/EPY
      DO 17 I=1,N1
      ALFA(I)=0.
      DO 17 J=1,N1
17    ALFA(I)=ALFA(I)+P(I,J)*BIGV(J)
      DEL=0.
      DO 18 I=1,NDIM
      DEL=DEL+G2(I)*(X2(I)-ALFA(I))
18    CONTINUE
      IF (ABS(DEL).GT.ETA(4)) GO TO 19
      IF (IFLAG.EQ.1) RETURN
      IFLAG=1
      GO TO 31
19    IFLAG=0
      DO 20 I=1,N1
      H(I)=X2(I)-ALFA(I)
      IF (DEL.GT.0) H(I)=-H(I)
20    CONTINUE
      DO 21 I=1,NDIM
      X1(I)=X2(I)
      G1(I)=G2(I)
21    CONTINUE
```

```
      F1=F2
      X1(N2)=X2(N2)
      X1(N1)=X2(N1)
      X2(N2)=ALFA(N2)
      X2(N1)=ALFA(N1)
      CALL MIN1D (FUNCT,X2,H,RO,NDIM,F2,G2,NUMF,IER,EPS,EST,RR,IPRINT)
      IF (NFAIL.EQ.1) RETURN
      IF (IER.NE.0) GO TO 30
      IF (DEL.GT.0) RO=-RO
      GG=0.
      DO 22 I=1,NDIM
      GG=GG+G2(I)*G2(I)
22    CONTINUE
      GG=SQRT(GG)
      KOUNT=KOUNT+1
      M=M+1
      IF (IPRINT.EQ.0) GO TO 23
      IF (MOD(KTB,IPRINT).NE.0) GO TO 23
      CALL OUTP (X2,F2,M,NDIM,GG,NUMF,RR)
23    CONTINUE
      KTB=KTB+1
      IF (MODE.EQ.2) GO TO 25
      IF (M.GT.MAX) GO TO 30
      NSOL=0
      DO 24 I=1,NDIM
      IF (ABS(RR(I)).GT.ETA(2)) NSOL=1
24    CONTINUE
      IF (NSOL.EQ.0) GO TO 26
      GO TO 29
25    IF ((GG.LT.ETA(1)).OR.(M.GT.MAX)) GO TO 26
      GO TO 29
26    CONTINUE
      IF (IPRINT.EQ.0) GO TO 27
      WRITE (6,33)
      CALL OUTP (X2,F2,M,NDIM,GG,NUMF,RR)
27    DO 28 I=1,NDIM
      X(I)=X2(I)
28    CONTINUE
      EST=F2
      NFAIL=0
      RETURN
29    CONTINUE
      IF (KOUNT.LE.N1) GO TO 11
      GO TO 10
30    CONTINUE
      IF (IPRINT.NE.0) PRINT 34, IER
      NFAIL=1
      RETURN
31    CONTINUE
      IF (IPRINT.NE.0) PRINT 35
```

```
      IER=0
      DO 32 I=1,NDIM
      X1(I)=X2(I)
      G1(I)=G2(I)
      H(I)=-G1(I)
32    CONTINUE
      F1=F2
      X1(N2)=X(N2)
      X1(N1)=X(N1)
      X2(N2)=X(N2)
      X2(N1)=X(N1)
      GO TO 3
C
C
C
33    FORMAT (*  SOLUTION FOUND*)
34    FORMAT (///,1X,*  THE OPTIMIZATION PROGRAM HAS FAILED--IER = *,I2)
35    FORMAT (///20X,*A RESTART HAS OCCURRED*///)
      END
```

```
      SUBROUTINE MINID (FUNCT,X,H,AMBDA,N,F,G,NUMF,IER,EPS,EST,RR,IPRINT
     1)
      COMMON /FAIL/ NFAIL
      DIMENSION H(1), X(1), G(1), RR(1)
      IER=0
      DY=0.
      HNRM=0.
      GNRM=0.
      DO 1 J=1,N
      HNRM=HNRM+ABS(H(J))
      GNRM=GNRM+ABS(G(J))
      DY=DY+H(J)*G(J)
1     CONTINUE
      IF (DY) 2,31,31
2     IF (HNRM/GNRM-EPS) 31,31,3
3     FY=F
      ALFA=2.*(EST-F)/DY
      IF (X(N+1).GT.0.) ALFA=X(N+1)*ALFA/2.
      AMBDA=1.
      IF (ALFA) 6,6,4
4     IF (ALFA-AMBDA) 5,6,6
5     AMBDA=ALFA
6     ALFA=0.
7     FX=FY
      DX=DY
      DO 8 I=1,N
      X(I)=X(I)+AMBDA*H(I)
8     CONTINUE
      CALL FUNCT (N,X,F,G,RR)
      IF (NFAIL.EQ.1) RETURN
      NUMF=NUMF+1
      IF (F.LT.FX) RETURN
      FY=F
      DY=0.
      DO 9 I=1,N
      DY=DY+G(I)*H(I)
9     CONTINUE
      IF (DY) 10,30,13
10    IF (FY-FX) 11,13,13
11    AMBDA=AMBDA+ALFA
      ALFA=AMBDA
      IF (HNRM*AMBDA-1.E10) 7,7,12
12    IER=2
      GO TO 31
13    T=0.
14    IF (AMBDA) 15,30,15
15    Z=3.*(FX-FY)/AMBDA+DX+DY
      ALFA=AMAX1(ABS(Z),ABS(DX),ABS(DY))
      DALFA=Z/ALFA
      DALFA=DALFA*DALFA-DX/ALFA*DY/ALFA
```

```
      IF (DALFA) 31,16,16
16    W=ALFA*SQRT(DALFA)
      ALFA=DY-DX+W+W
      IF (ALFA) 17,18,17
17    ALFA=(DY-Z+W)/ALFA
      GO TO 19
18    ALFA=(Z+DY-W)/(Z+DX+Z+DY)
19    ALFA=ALFA*AMBDA
      DO 20 I=1,N
      X(I)=X(I)+(T-ALFA)*H(I)
20    CONTINUE
      CALL FUNCT (N,X,F,G,RR)
      IF (NFAIL.EQ.1) RETURN
      NUMF=NUMF+1
      IF (F.LT.FX) GO TO 30
      IF (F-FX) 21,21,22
21    IF (F-FY) 30,30,22
22    DALFA=0.
      DO 23 I=1,N
      DALFA=DALFA+G(I)*H(I)
23    CONTINUE
      IF (DALFA) 24,27,27
24    IF (F-FX) 26,25,27
25    IF (DX-DALFA) 26,30,26
26    FX=F
      DX=DALFA
      T=ALFA
      AMBDA=ALFA
      GO TO 14
27    IF (FY-F) 29,28,29
28    IF (DY-DALFA) 29,30,29
29    FY=F,
      DY=DALFA
      AMBDA=AMBDA-ALFA
      GO TO 13
30    AMBDA=AMBDA-ALFA
      RETURN
31    CONTINUE
      IF (DY.GE.0.) IER=-2
      IF (GNRM.LE.1.E-10) GO TO 32
      IF (HNRM/GNRM.LE.EPS) IER=-3
32    CONTINUE
      IF (DALFA.LT.0.) IER=-1
      IF (IPRINT.NE.0) PRINT 33, IER
      NFAIL=1
      RETURN
C
C
C
33    FORMAT (///10X,* ERROR HAS OCCURRED, IER=*,I2,///)
      END
```

```
      SUBROUTINE OUTP (XNEW,FQ,KOUNT,N1,GG,NUMF,R)
      DIMENSION XNEW(1), R(1)
      WRITE (6,6) KOUNT,NUMF,GG,FQ,(XNEW(I),I=1,4),(R(I),I=1,4)
      IF (N1.LT.4) RETURN
      NN=N1-3
      GO TO (1,2,3,4,5), NN
1     RETURN
2     WRITE (6,7) XNEW(5),R(5)
      RETURN
3     WRITE (6,8) (XNEW(I),I=5,6),(R(I),I=5,6)
      RETURN
4     WRITE (6,9) (XNEW(I),I=5,7),(R(I),I=5,7)
      RETURN
5     WRITE (6,10) (XNEW(I),I=5,8),(R(I),I=5,8)
      RETURN
C
C
C
6     FORMAT (1X,I3,I4,6E14.5,4E11.3)
7     FORMAT (36X,E14.5,42X,E11.3)
8     FORMAT (36X,2E14.5,28X,2E11.3)
9     FORMAT (36X,3E14.5,14X,3E11.3)
10    FORMAT (36X,4E14.5,4E11.3)
      END
```

SUBROUTINE MOMENT(CM,N,CC,M)

Purpose

To provide an estimate of one or more (M) of the first four moments of y, where

$$y = g(x_1,x_2,\ldots,x_n)$$

and the first M moments of the x's are known.

Method

The first M moments of the first four moments of y are estimated in terms of the moments of the x_i's by using a truncated Taylor's series expansion. The user must supply an evaluation of $g(x_1,x_2,\ldots,x_n)$, $\frac{\partial g}{\partial x_i}$ and $\frac{\partial^2 g}{\partial x_i^2}$.

Input Variables

N	number of independent variables.
M	number of moments required. Note that $M \leq 4$.
CM(I,J)	array containing the first M moments of the independent variables, dimensioned (N,M)

Output Variables

CC(I)	array containing the values of the first M moments dimensioned with the value of M.

Programing Information

The calling program must provide dimensioning as given above. The user must define the function $g(x_1,x_2,\ldots,x_n)$, and the first and second

partial derivatives. See subroutine DERV.

Listing

The following listing is for subroutine MOMENT.

```
      SUBROUTINE MOMENT(CM,N,CC,M)
      DIMENSION CM(N,M),CC(4) ,DF1(4),DF2(4)
      DO 6 I=1,N
      CC(I)=CM(I,1)
6     CONTINUE
      CALL DERV (FUN,DF1,DF2,N,CC)
      CC(1)=FUN
      DO 7 I=2,4
      CC(I)=0.0
7     CONTINUE
      DO 10 I=1,N
      CC(1)=CC(1)+.5*(DF2(I)*CM(I,2))
      IF(M.EQ.1) GO TO 10
      CC(2)=CC(2)+DF1(I)**2*CM(I,2)+DF1(I)*DF2(I)*CM(I,3)
      IF(M.EQ.2) GO TO 10
      CC(3)=CC(3)+DF1(I)**3*CM(I,3)
      IF(M.EQ.3) GO TO 10
      SUM=0.0
      KJ=I+1
      IF (KJ.GT.N) GO TO 9
      DO 8 J=KJ,N
      SUM=SUM+6.*(DF1(I)*DF1(J))**2*CM(I,2)*CM(J,2)
8     CONTINUE
9     CONTINUE
      CC(4)=CC(4)+SUM+DF1(I)**4*CM(I,4)
10    CONTINUE
      RETURN
      END
```