BU-289-M

IDENTIFIABILITY AND STRONGLY CONSISTENT ESTIMATES FOR FINITE MIXTURES OF DISCRETE DISTRIBUTIONS

Keewhan Choi
Cornell University

ABSTRACT

The problems with which we are concerned in this paper are those of identifiability and strongly consistent estimates for a mixing measure of a finite mixture of discrete (and finite) distribution functions. We present an elementary proof for the fact that the identifiability is a necessary and sufficient condition for the existence of a strongly consistent estimate of the mixing measure. Several strongly consistent estimates of the mixing measure of finite and discrete distributions are proposed and their strong consistency is proven. Results of a small Monte Carlo study of the sampling distributions of the various estimates are given.

Paper No. BU-120, Biometrics Unit, and Paper No. 500, Plant Breeding Department, Cornell University

IDENTIFIABILITY AND STRONGLY CONSISTENT ESTIMATES FOR FINITE MIXTURES OF DISCRETE DISTRIBUTIONS

Keewhan Choi
Cornell University

Summary

The problems with which we are concerned in this paper are those of identifiability and strongly consistent estimates for a mixing measure of a finite mixture of discrete (and finite) distribution functions. We present an elementary proof for the fact that the identifiability is a necessary and sufficient condition for the existence of a strongly consistent estimate of the mixing measure. Several strongly consistent estimates of the mixing measure of finite and discrete distributions are proposed and their strong consistency is proven. Results of a small Monte Carlo study of the sampling distributions of the various estimates are given.

1. Introduction

Mixtures of distribution functions are of considerable interest not only for their mathematical aspects but also for the large number of applied problems in which mixtures occur. Estimation of mixing measures of known component distributions is of the most general interest, however, it is necessary to investigate first the identifiability of mixing measures.

In a series of recent papers, Teicher (1960, 1961, 1963) has investigated extensively the identifiability problem. In particular, Teicher (1963) has

---

Paper No. BU-120, Biometrics Unit, and Paper No. 500, Plant Breeding Department, Cornell University

given a necessary and sufficient condition for the identifiability of finite mixtures. In Section 2, we present an elementary proof for the fact that identifiability is equivalent to the existence of a strongly consistent (i.e. converging to the true parameter with probability one) estimate for the mixing measure of finite and discrete distributions.

Pearson (1894), Rao (1952), Rider (1961a, 1961b) and Blischke (1962, 1964) have considered the estimation problem for the parameters of the component distribution functions comprising a finite mixture and the mixing measure. All the authors have used the method of moments to obtain the estimates for a mixture of finitely many members of a parametric family of distribution functions. (Blischke (1964) has discussed the other methods also.) In Section 3 we propose several estimates for the mixing measure. We prove that the proposed estimates are strongly consistent, i.e., they converge with probability one to the true measure. In Section 4, results of a small Monte Carlo study are given to indicate the sampling distributions of the estimates proposed in Section 3.

For extensive discussion with references on applications of mixtures of distributions see Blischke (1963).

2. Equivalence of identifiability and existence of a strongly consistent estimate of the mixing measure

Let $f = \{f_{i\alpha}, \alpha = 1,2,3,\ldots,m\}$ be a known family of m discrete one-dimensional distribution functions for $i = 1,2,\ldots,r$ where $f_{ij} = \Pr\{X = i | \alpha = j\}$. Let $G = (g_1,g_2,\ldots,g_m)^t$ be any column vector of positive real numbers whose sum is one. Then the new distribution function

$$\Pr\{X = i\} \equiv P_G(i) = \sum_{\alpha=1}^{m} f_{i\alpha} g_\alpha \qquad \text{for } i = 1,2,\ldots,r$$

is called a G-mixture of f and G the mixing measure. If we let F denote the matrix $(f_{i\alpha})$, $i = 1,\ldots,r$, $\alpha = 1,2,\ldots,m$, then $P_G(i)$ the i^{th} element of $P_G \equiv F \cdot G = (P_G(1), P_G(2), \ldots, P_G(r))^t$.

The problem is to find a strongly consistent estimate of G from n independent observations $x_1, x_2, \ldots, x_n$ with the common distribution P_G. However, we must first investigate the question of identifiability.

Let $\mathcal{G}$ denote the class of all such discrete mixing measures and $\mathfrak{F}$ the induced class of mixtures. Then $\mathfrak{F}$ is said to be identifiable in $\mathcal{G}$ (with respect to f) if G and $\overline{G}$ are any two mixing measures such that $F \cdot G = F \cdot \overline{G}$ then $G = \overline{G}$. (i.e. columns of F are linearly independent.) For the sake of brevity let us define, in this section, estimability to mean the existence of a strongly consistent estimate of the unknown G.

<u>Theorem 1</u>. Identifiability (i.e. $\mathfrak{F}$ is identifiable) is a necessary and sufficient condition for estimability.

<u>Proof</u>. Necessity is obvious. Sufficiency will be proven in the following two steps.

(i) Identifiability $\Leftrightarrow$ $m = \text{rank}(F)$.

(ii) $m = \text{rank}(F) \Rightarrow$ estimability.

The proof of (i) is immediate from the definition of identifiability and the fact that the m-dimensional Euclidean space cannot have more than m linearly independent vectors.

The proof of (ii) depends on the result in Section 7 of Robbins (1964) that

Estimability is equivalent to the condition (R):

(R) If $G = \bar{G}$ are any two probability vectors such that for every set B $\sum_{\alpha=1}^{m} g_\alpha Q_\alpha(B) = \sum_{\alpha=1}^{m} \bar{g}_\alpha Q_\alpha(B)$, then $G = \bar{G}$, where

$$Q_k(B) \equiv \Pr\{X\epsilon B|\alpha=k\} = \sum_{i\epsilon B} f_{ik} .$$

From the definition of identifiability it follows that:

if $\sum_{\alpha=1}^{m} x_\alpha F_\alpha(k) = 0$ for all $k(= 1,2,\ldots,r)$ where $F_\alpha(k) = \sum_{i\leq k} f_{i\alpha}$,

then $x_\alpha = 0$ for all $\alpha(= 1,2,\ldots,m)$. This implies:

if $\sum_{\alpha=1}^{m} x_\alpha [F_\alpha(k) - F_\alpha(k')] = 0$ for all k, k' such that $k > k'$,

then $x_\alpha = 0$ for all α, which implies (R).

Hence, Identifiability $\Rightarrow$ Estimability.

The theorem could be deduced from Theorem 1 of Teicher (1963) and Section 7 of Robbins (1964). Nevertheless we have presented an elementary proof for completeness.

3. Strongly consistent estimates of the mixing measure

Let F_n be the emperical distribution of $(x_1,x_2,\ldots,x_n)$, or equivalently let $n_k(k = 1,2,\ldots,r)$ denote the number of observations x_j which are equal to k. Then a sequence of vectors $G^*(n) = \{g^*_{j(n)} : j = 1,2,\ldots,m\}$ which minimizes

$$(E1) = \sum_{i=1}^{r} (P_G(i) - n_i/n)^2 P_G(i) \qquad \text{(Ch estimate)}$$

is a strongly consistent estimate of the true mixture G. Since we are assuming that every component of the true mixing measure G is positive we exclude from our consideration those $G^*(n)$ every component of which approaches zero as n increases.

Let $\delta(P_G, F_n)$ denote the expression (E1). Then, by definition

(1) $\delta(P_{G^*(n)}, F_n) = \inf_G \delta(P_G, F_n)$.

By the Glivenko-Cantelli theorem we have with probability one

$$n_i/n \to P_G(i) \quad \text{for all } i .$$

Hence by the Slutsky theorem [p 255 Cramér (1946)]

(2) $\delta(P_G, F_n) \to 0$

with probability one as $n \to \infty$.

Hence by (1)

(3) $\delta(P_{G^*(n)}, F_n) \to 0$ with probability one.

$$\begin{aligned}\delta(P_{G^*(n)}, F_n) &= \sum_{i=1}^{r} (P_{G^*(n)}(i) - n_i/n)^2 P_{G^*(n)}(i) \\ &= \sum_{i=1}^{r} (P_{G^*(n)}(i) - P_G(i) + P_G(i) - n_i/n)^2 P_{G^*(n)}(i) \\ &= \sum_{i=1}^{r} [(P_{G^*(n)}(i) - P_G(i))^2 + (P_G(i) - n_i/n)^2 \\ &\qquad + 2(P_{G^*(n)}(i) - P_G(i))(P_G(i) - n_i/n)] P_{G^*(n)}(i)\end{aligned}$$

Since $n_i/n \to P_G(i)$ with probability one, by the Slutsky theorem again,

(4) $\delta(P_{G^*(n)}, F_n) \to \sum_{i=1}^{r} (P_{G^*(n)}(i) - P_G(i))^2 P_{G^*(n)}(i)$.

Since the expression on left hand side of (4) approaches zero as n increases, so does the one on the right hand side.

Since each term of the expression on the right hand side of (4) is non-negative, each term must approach zero. To wit, with probability one,

(5) $(P_{\overset{*}{G}(n)}(i) - P_G(i))^2 P_{\overset{*}{G}(n)}(i) \to 0$ for all i,

which implies

(6) $P_{\overset{*}{G}(n)}(i) \to P_G(i)$ for all i.

To see the validity of the implication, it is sufficient to consider the following case only.

For all i, $g_i(n)$ converges and

there exists k such that $g_k(n)$ converges to a positive number.

Then the proof is immediate from the inspection of (5) (replacing $P_{\overset{*}{G}(n)}(i)$ by $\sum_{j=1}^{m} g_{j(n)} f_{ij}$) .

Now (6) implies:

(7) $\overset{*}{g}_{j(n)} \to g_i$ for all j.

In matrix notation (6) states

$$\lim_{n\to\infty} F(\overset{*}{G}(n) - G) = 0$$

which is equivalent to

$$F \lim_{n\to\infty} (\overset{*}{G}(n) - G) = 0.$$

Since the rank of F is m

$\lim_{n\to\infty} (\overset{*}{G}(n) - G) = 0$, which is equivalent to (7).

In the same manner, strong consistency can be easily proven for the estimate G(n) which minimizes any one of the following expressions:

(E2) $\sum_{i=1}^{r} (P_G(i) - n_i/n)^2 n_i/n$ (modified Ch estimate)

(E3) $\sum_{i=1}^{r} (P_G(i) - n_i/n)^2/P_G(i)$ (minimum χ^2 like estimate)

(E4) $\sum_{i=1}^{r} (P_G(i) - n_i/n)^2 n/n_i$ (modified minimum χ^2 like estimate)

(E5) $\sum_{i=1}^{r} (P_G(i) - n_i/n)^2$ (least square estimate) .

A quite different kind of estimator of the mixing measures has been proposed by Robbins (1964). According to Robbins' method, for instance, g_1 is estimated as the (normalized) orthogonal complement of the projection of the first column of F onto the space spanned by all but the first column of F. Robbins' estimate is also strongly consistent. It is also obvious from Robbins (1964) that the asymptotic distribution of Robbins' estimate is (multivariate) normal.

4. Monte Carlo study of the various estimates for the mixing measure

We are currently investigating the asymptotic distributions of the estimates proposed in Section 3. In this section the results of a small Monte Carlo study of the sampling distributions of the estimates discussed in Section 3 are given. For comparison, the mixing measures are estimated also by Robbins' method. Using pseudo-random numbers sets of independent observations are generated from a (g_1,g_2,g_3) mixture of three Binomial distributions with the same n, and different p's. Then the various estimates are computed. The values of n g_i's and p_i's used in generating the observations are:

$$n = 8, \quad p_1 = \frac{1}{2}, p_2 = \frac{1}{3}, p_3 = \frac{1}{6}$$

$$g_1 = g_2 = g_3 = \frac{1}{3} .$$

In Table 1, the results from 300 samples of 32 observations each are summarized: means and variances of the various estimates of (g_1, g_2, g_3) are presented. Given in Figures 1 through 6 are the sampling distributions of the various estimates of g_1. It is very difficult to select any one of the estimates from studying Table 1 and the figures. If we use the sum of the variances as a criterion, the minimum χ^2 like estimate seems to be the best. However, a much larger Monte Carlo study is required before any one of the estimates can be chosen as the best.

TABLE 1 (True value of $g_1 = g_2 = g_3 = \frac{1}{3}$)

	g_1	g_2	g_3	Sum of the variances
Robbins' estimates				
Mean	.3775	.3318	.3603	
Variance	.0650	.1063	.0411	.2124
Ch estimates (E1)				
Mean	.3522	.3129	.3349	
Variance	.0391	.0932	.0342	.1665
Modified Ch estimates (E2)				
Mean	.3235	.3525	.3239	
Variance	.0475	.0969	.0340	.1784
Minimum χ^2 like estimates (E3)				
Mean	.3958	.2608	.3435	
Variance	.0359	.0695	.0237	.1291

Modified minimum χ^2 like estimates (E4)

Mean	.2589	.4120	.3259	
Variance	.0459	.0961	.0372	.1792

Least square estimates (E5)

Mean	.3471	.3159	.3463	
Variance	.0401	.0850	.0332	.1583

FIGURE 1

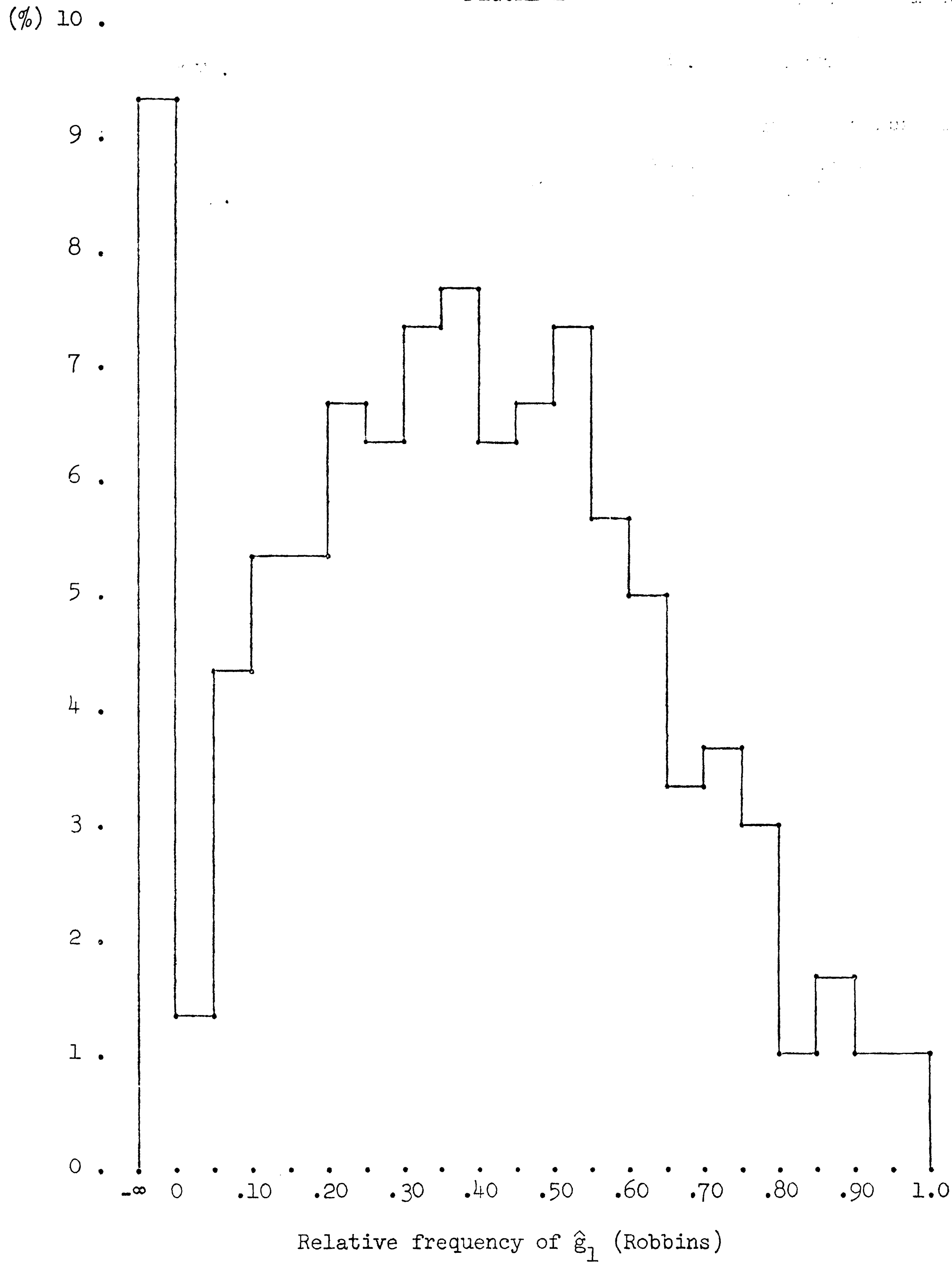
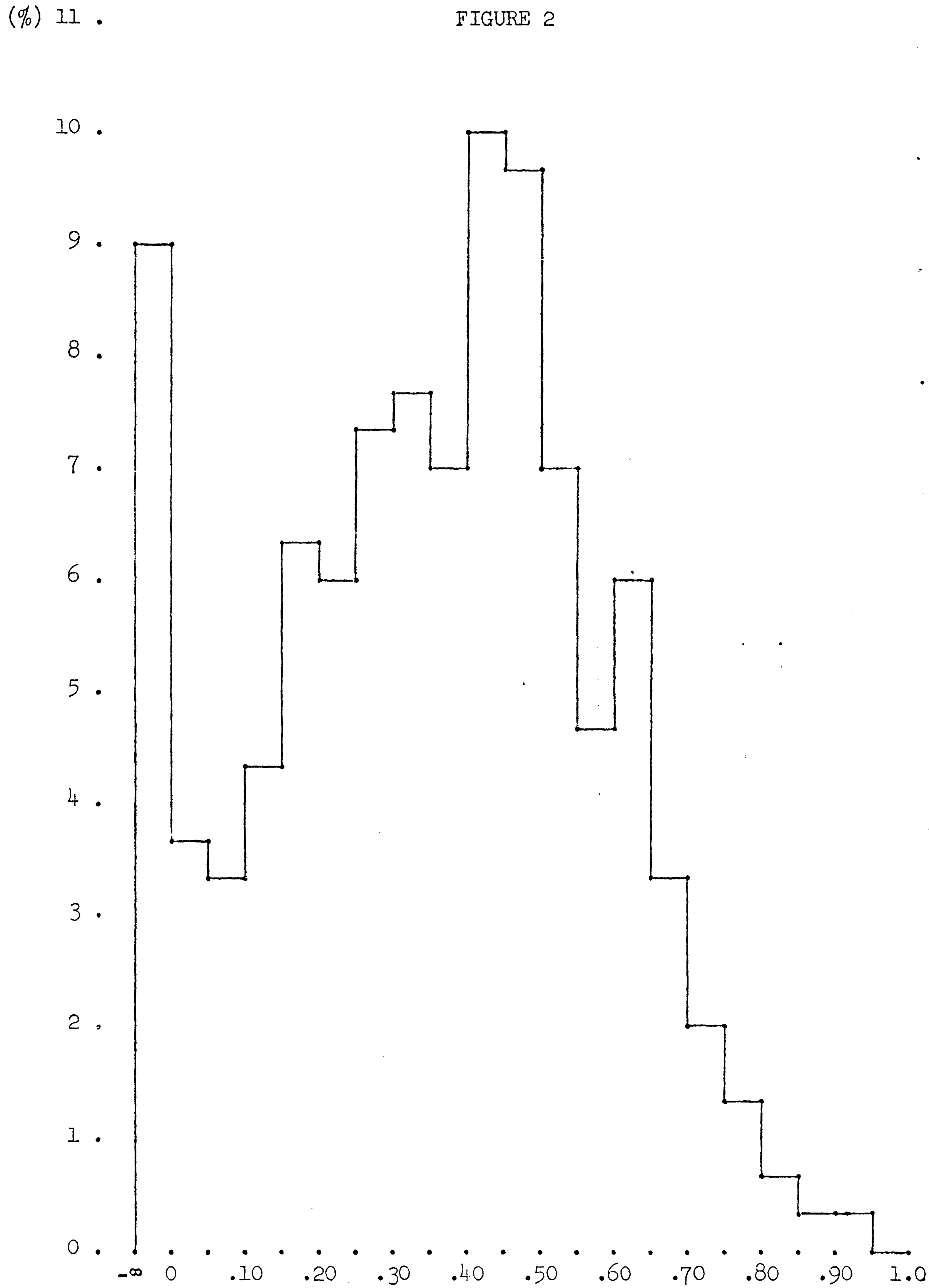

(%) 10
9
8
7
6
5
4
3
2
1
0
-∞
0
.10
.20
.30
.40
.50
.60
.70
.80
.90
1.0
Relative frequency of $\hat{g}_1$ (Robbins)

FIGURE 2



Relative frequency of $\hat{g}_1$ (Ch) E1

Relative frequency of $\hat{g}_1$ (modified Ch) E2

FIGURE 3

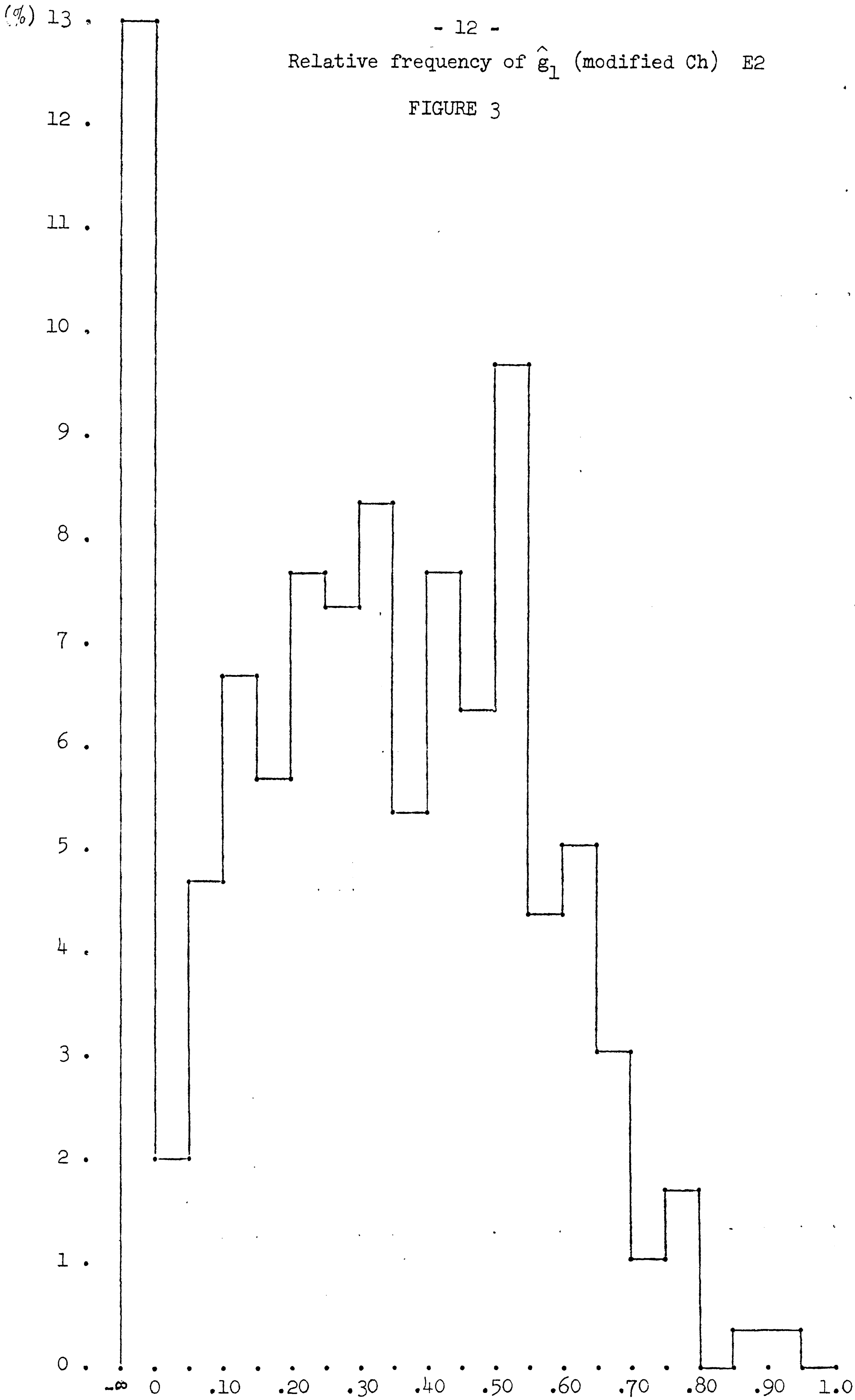
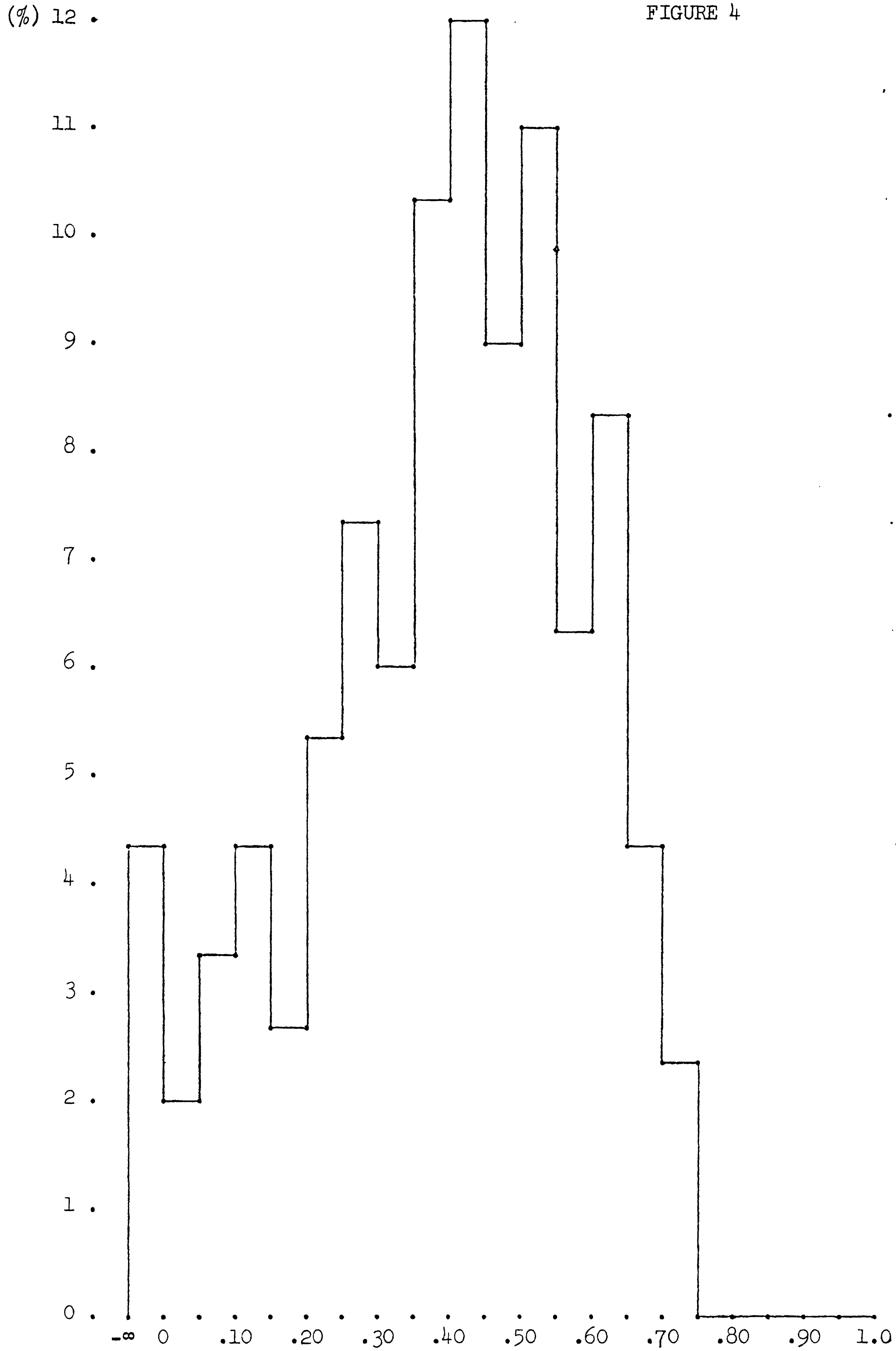
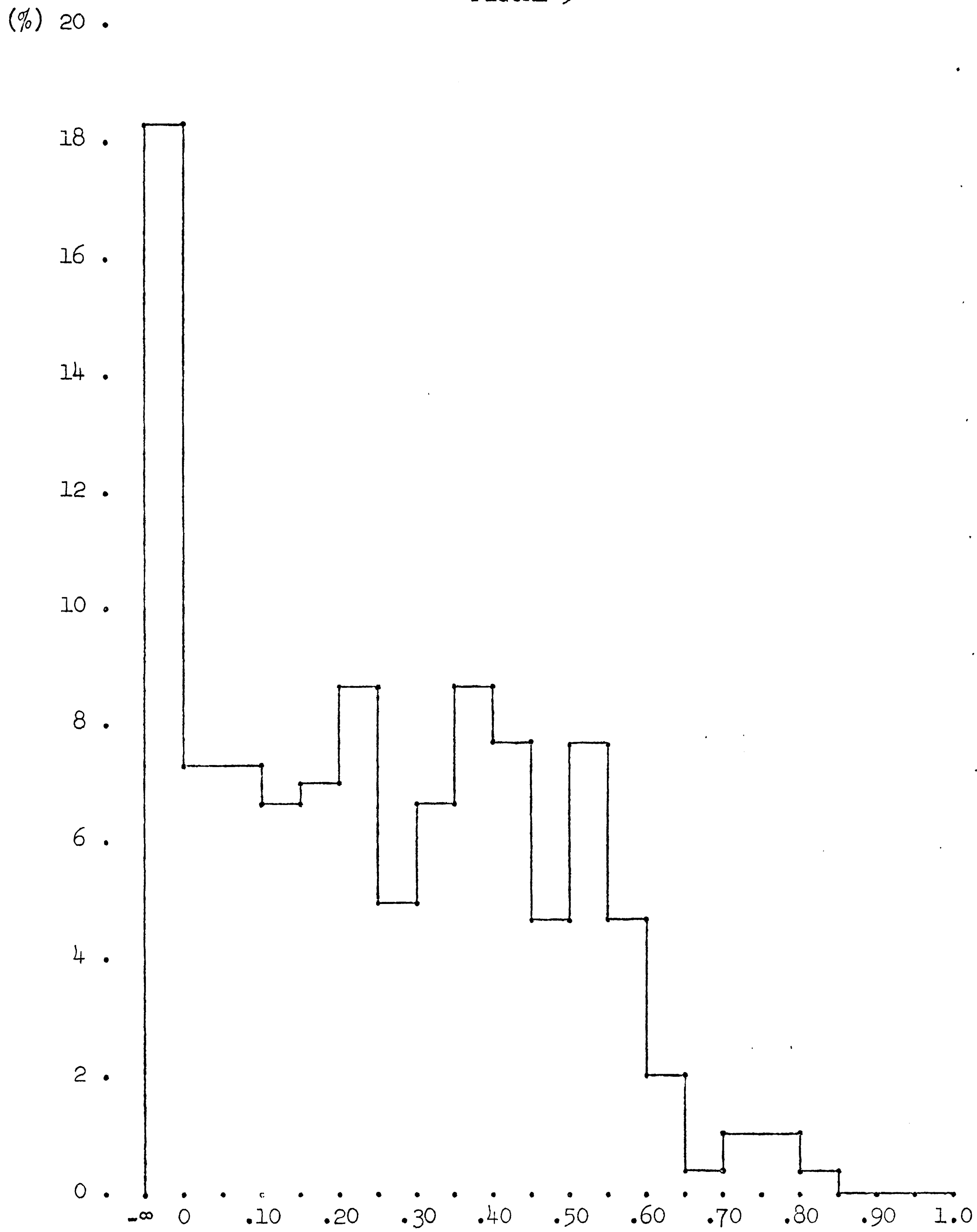


FIGURE 4



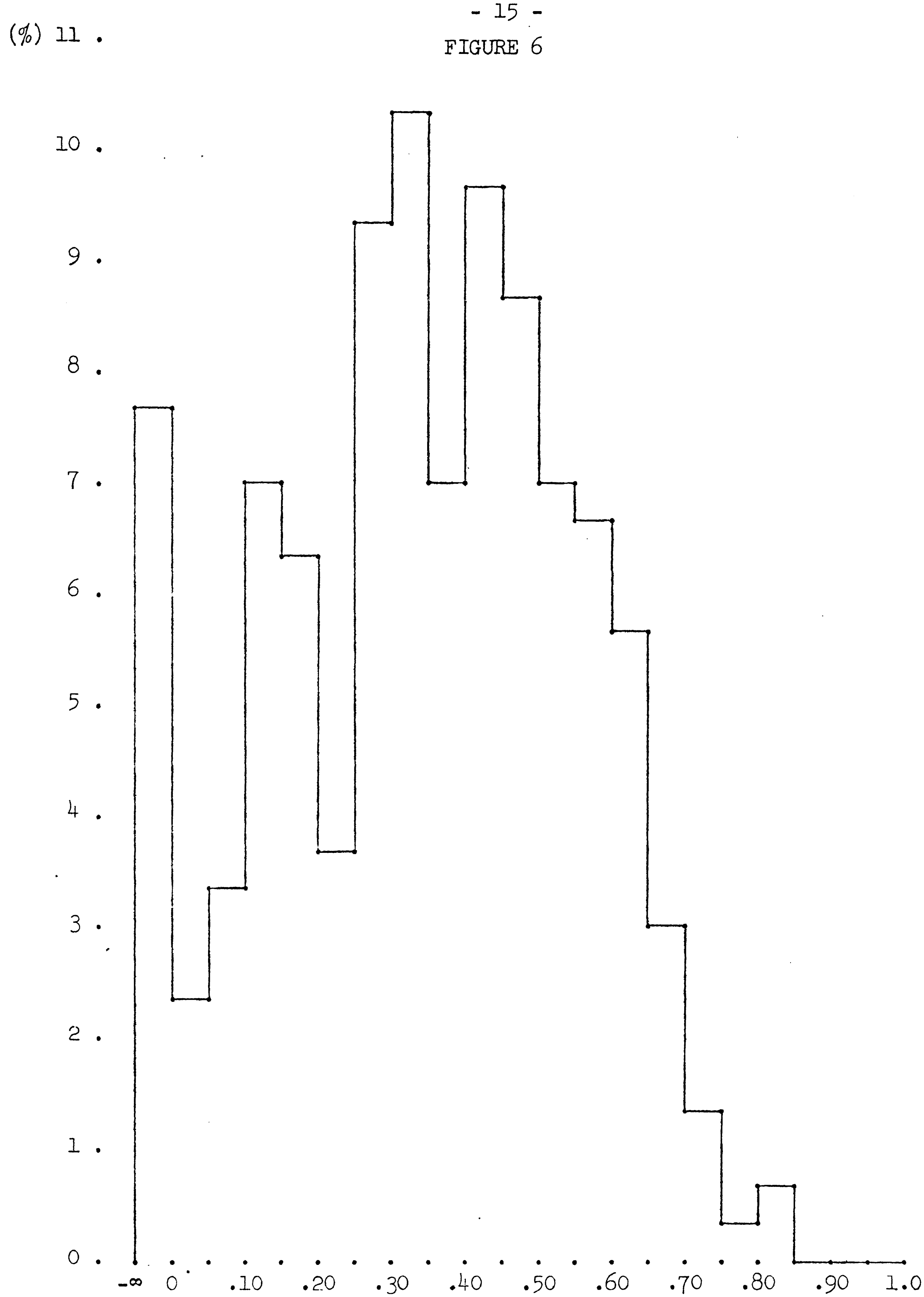
Relative frequency of $\hat{g}_1$ (minimum χ^2 like) E3

FIGURE 5



Relative frequency of $\hat{g}_1$ (modified minimum χ^2 like) E4

FIGURE 6



Relative frequency of $\hat{g}_1$ (Least Square) E5

Acknowledgements

The author wishes to thank Professors J. Wolfowitz and H. E. Schaffer for several helpful discussions. Thanks are also due to Mrs. H. E. Schaffer for programming the computations in Section 4. The proof in Section 3 was inspired by the minimum distance method of Wolfowitz (1957). This research was supported in part by the National Institute of Health Research Grant PHG-GM-05900.

References

Blischke, W. R. (1962). Moment estimators for the parameters of a mixture of two binomial distributions. Ann. Math. Statist. 33 444-454.

Blischke, W. R. (1963). Mixtures of discrete distributions. Paper presented at the International Symposium on the Classical Discrete Distributions. McGill University, Montreal, August 15-20, 1963.

Blischke, W. R. (1964). Estimating the parameters of mixtures of binomial distributions. J. Amer. Statist. Assoc. 59 510-528.

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. Phil. Trans. Roy. Soc. 185A 71-110.

Rao, C. R. (1952). Advanced Statistical Methods in Biometric Research. Wiley, New York.

Rider, P. R. (1961a). The method of moments applied to a mixture of two exponential distributions. Ann. Math. Stat. 32 143-147.

Rider, P. R. (1961b). Estimating the parameters of mixed Poisson, binomial, and Weibull distributions by the method of moments. Bull. Inst. Int. Stat. 39 (Part 2).

Robbins, H. (1964). The emperical Bayes approach to statistical decision problems. Ann. Math. Stat. 35 1-20.

Teicher, H. (1960). On the mixture of distributions. Ann. Math. Stat. 19 360-369.

Teicher, H. (1961). Identifiability of mixtures. Ann. Math. Stat. 32 244-248.

Teicher, H. (1963). Identifiability of finite mixtures. Ann. Math. Stat. 34 1265-1269.

Wolfowitz, J. (1957). The minimum distance method. Ann. Math. Stat. 38 75-88.