DOCUMENT RESUME

ED 247 271 TM 840 445

AUTHOR Peng, Chao-Ying Joanne
TITLE Utility of the Beta-Multinomial Distribution in Multiple Classification Scheme.
PUB DATE Apr 84
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Classification; *Criterion Referenced Tests; Mastery Tests; *Mathematical Models; *Statistical Distributions; Student Evaluation
IDENTIFIERS Beta Binomial Test Model; *Beta Multinomial Test Model

ABSTRACT
This study is an attempt to answer the following research question: can the reliability of a criterion-referenced test be accurately determined according to a multiple classification of the student's performance? Specifically, the study pursues the beta-multinomial model, which postulates the probability distribution of an examinee's degree of mastery on a criterion-referenced test. From this model, a procedure for assessing the reliability of the testing instrument was developed. Simulated data based on the beta-multinomial distributions did not depart significantly from those generated by the beta-binomial model. However, these results should not preclude the utility of beta-multinomial models in this context. (Author/BW)

ED247271

UTILITY OF THE BETA-MULTINOMIAL DISTRIBUTION IN MULTIPLE CLASSIFICATION SCHEME

Chao-Ying Joanne Peng
Dept. of Counseling and Educational Psychology
Indiana University at
Bloomington

Paper presented at the annual meeting of the
American Educational Research Association
April 23-27, 1984
New Orleans, U.S.A.

TM 810 445

UTILITY OF THE BETA-MULTINOMIAL DISTRIBUTION IN MULTIPLE CLASSIFICATION SCHEME

Abstract

The present study is an attempt to answer the following research question: Can the reliability of a criterion-referenced test be accurately determined according to a multiple classification of the student's performance? Specifically, this study pursues a sound statistical model, i.e., the beta-multinomial model, which postulates the probability distribution of examinee's degree of mastery on a criterion-referenced test. From this model, a procedure for assessing the reliability of the testing instrument can then be developed. Ideally and finally, several real-life data sets should have been employed in order to justify empirically (or refine) this reliability estimation procedure. Results from this study should and would solve some knotty psychometric difficulties which are presently hindering the progress of the criterion-referenced testing movement.

Background

Within the domain of criterion-referenced testing, various methods have existed in the literature which are intended to assess the reliability of a test (Subkoviak, 1979). Among these procedures, Huynh's single-administration approach has received much attention due to the elegance of its model and tolerable bias associated with its estimates (Huynh, 1976; Subkoviak, 1978). Subsequently, Huynh's procedure was well investigated and simplified for classroom teachers or practitoners who might not have access to a computer (Peng and Subkoviak, 1980).

The Beta-Binomial Model

Two major assumptions underlie Huynh's procedure:

(I) A binomial density function is assumed for the distribution of scores (x) for an examinee with true ability ξ over repeated n-item tests. Therefore,

$$f(x|\xi) = \binom{n}{x} \xi^x (1-\xi)^{n-x} \quad , x = 0, \cdots, n.$$

ξ is the proportion of items in the item population that an examinee can correctly answer.

(II) A beta distribution for ξ is assum[illegible] the population.

Under these assumptions, it can be shown ([illegible] and Lord, 1962) that the probability distribution of x is a beta-binomial (or negative hyper-geometric) distribution with the following form:

$$f(x)=\binom{n}{x}B(\alpha+x,\ n+\beta-x)/B(\alpha,\beta),$$

where n= number of items on a test and

B(,)= a beta function defined by the parameters in the parenthesis.

A bivariate beta-binomial distribution is determined similarly,

$$f(x,y)=\frac{\binom{n}{x}\binom{n}{y}}{B(\alpha,\beta)}B(\alpha+x+y,\ 2n+\beta-x-y).$$

Reliability Indices derived from the Beta-Binomial Model

Under the beta-binomial model, a criterion-referenced test is simply a mastery test. A mastery test typically classifies an examinee into one of the two categories: master or nonmaster, according to a predetermined criterion or cutoff. Figure 1 below depicts this general decision-making framework.

		Form Y: Master	c	Form Y: Nonmaster	Marginal Proportions
Form X	Master	P_{00}			P_0
	c				
	Nonmaster			P_{11}	P_1
		P_0		P_1	

When two parallel forms X and Y exist , the probability of consistent classification of pupils is composed of two elements: the probability of a nonmaster consistently identified by both forms and that of a master again by X and Y. Mathematically, this probability can be expressed as

$$P_{\text{consistent classification}} = P = P_{00} + P_{11}.$$

This binary classification is equally imposed on individual items from the perspective of a beta-binomial model. Alternatively, a standardized kappa coefficient can be used also to suffice the purpose of quantifying a reliability . This leads into the following definition

$$\text{Kappa} = \frac{P - P_{\text{chance}}}{\max(P) - P_{\text{chance}}} = \frac{P - (P_0^2 + P_1^2)}{1 - (P_0^2 + P_1^2)}.$$

Statement of the Problem

Unfortunately, the Huynh's approach as well as the simplied procedure assumes that an examinee either masters or fails a test. In order words, these approaches are restricted to mastery tests only. It is ,however, more realistic to assume that a typical pupil is capable of mastering a portion, if not the entirety, of all the materials taught. Hence, a multiple classification scheme on items and tests seems reasonable for determining a student's level of mastery on a criterion-referenced test. This suggests the development of the beta-multinomial model, which is an expansion of the beta-binomial model underlying the Huynh method.

The Beta-Multinomial Model

Three useful references are given by Cheng (1964, in modern Chinese), Ishii and Hayakawa (1960) and Mosimann (1962). The original manuscripts were published in separate and yet remote locations around the world; hence, they singalled an alarmingmessage for more headaches in days to come as long as I remained interested in pursuing this line of research (Sigh!)

Two major assumptions implied by the beta-multinomial model:

(1) A multinomial density function is assumed for the conditional distribution of scores x ($=x_1 + w \cdot x_2$) for an examinee with true ability ξ ($= \xi_1 + \omega \cdot \xi_2$) over repeated N-items test.

$$f\,(X=X_1+wX_2 \mid \zeta=\zeta_1+w\zeta_2)$$

$$= \frac{N!}{X_1!\,X_2!\,(N-X_1-X_2)!}\,(\zeta_1)^{X_1}\,(\zeta_2)^{X_2}\,(1-\zeta_1-\zeta_2)^{N-X_1-X_2},$$

$$0 \le (X=X_1+wX_2) \le N$$

where X_1=# of items that an examinee can completely master,

X_2=# of items that an examinee can partially master,

w =partial credit awarded to items on which an examinee demonstrates partial mastery which equals a constant term in the equation,

ζ_1=the proportion of items in the item population that an examinee can correctly answer, and

ζ_2=the proportion of items in the item population that an examinee can partially answer.

(2) A multivariate beta distribution for ζ_1 & ζ_2 is assumed across the population of examinees.

Under these assumptions, it can be shown (Mosimann, 1962) that the probability distribution of X is a compound beta-multinomial

$$f(X) = \frac{N!}{X_1!\,X_2!\,(N-X_1-X_2)!}\;\frac{\int_0^1 d\xi_1 \int_0^{1-\xi_1} (\xi_1)^{x_1+\alpha_1-1} (\xi_2)^{x_2+\alpha_2-1} d\xi_2}{B(\alpha_1\,,\sum_{i=2}^{3}\alpha_i)\,B(\alpha_2\,,\alpha_3)}$$

where B(,)= a beta function defined by the parameters in the parentheses.

Estimation procedures of this complex parametric model are provided in Cheng (1964). However, Cheng's procedures are far too sophisticated to be implemented by practitioners in education. Simplified procedures (such as the method of moments) ought to be developed and also the applications of the beta-multinomial model in the literature deserves an in-depth review.

When the beta-multinomial model is generalized to a joint distribution of scores x and y on parallel tests, a bivariate beta-multinomial distribution should result (by mathematical derivation). This bivariate distribution, denoted by f(x,y), should have the same set of parameters as f(x), since x and y are obtained from parallel tests and identical criteria should be enforced in both cases. Hence, estimated parameters developed in any estimation procedure should be sufficient in determining the bivariate distribution of scores, f(x,y), which would result if two tests were indeed administered. This rationale constitutes a sound basis for developing a single-administration approach in assessing the reliability of a criterion-referenced test.

Proposed Procedure for Assessing Reliability based on Beta-Multinomial Model

Two phases: { Simulated Data ; and
Real Data (very difficult to locate)

Simulated Data. Four steps are necessary :

Step 1-- Various values of alphas are considered according to the specification in Table 1 (page 6).

Step 2-- Specifications on test length (N) and cutoff scores (C_1 and C_2) are included in Table 2 (page 7).

Step 3-- Generate the f(x) and f(x,y) distributions based on Steps 1 and 2.

Step 4-- Develop a single-administration approach to compute P or kappa. Tentatively, $P=P_{00} + P_{11} + P_{22}$ and

$$\text{Kappa} = \frac{P - P_{chance}}{1 - P_{chance}}.$$

Table 1

Selected Beta Distributions for Study

Case	α_1	α_2	α_3	general description
I.	1	1	1	Uniform
II	.5	.5	.5	U-shaped
III	2	2	2	Symmetric, unimodal & platykurtic
IV	3	3	3	Symmetric, unimodal & leptokurtic
V	6	2	2	Negatively skewed

Table 2

SELECTED VALUES OF N, C_1 and C_2

N	C_1			C_2		
	45%	55%	65%	75%	85%	95%
5	3 (2.25)	3 (2.75)	4 (3.25)	4 (3.75)	5 (4.25)	5 (4.75)
10	5 (4.50)	6 (5.50)	7 (6.50)	8 (7.50)	9 (8.50)	10 (9.50)
15	7 (6.75)	9 (8.25)	10 (9.75)	12 (11.25)	13 (12.75)	15 (14.25)
20	9 (9.0)	11 (11.0)	13 (13.0)	15 (15.0)	17 (17.0)	19 (19.0)
30	14 (13.5)	17 (16.5)	20 (19.5)	23 (22.5)	26 (25.5)	29 (28.5)

<u>Real Data Analysis</u> . Also four steps are to be executed:

Step 1-- Estimate α_1, α_2, and α_3 via the method of moments. This needs an in-depth review of the literature).

Step 2-- Generate the f(x) and f(x,y) distribution based on Step 1 above.

Step 3-- Compute $\hat{P}$ and $\widehat{kappa}$ according to the single-administration procedure developed in Step 4 under the simulated study.

Step 4-- Compare $\hat{P}$ against true P obtained from the test-retest results;

Also, perform the same contrast between $\widehat{kappa}$ and true kappa to determine whether the beta-multinomial model along with the single test administration procedure yield satisfactory results.

Preliminary Results Obtained from Simulated Data

In simulating artificial data from the beta-multinomial model; the actual criterion scores (Table 2) were never utilized. Instead, the proportion of mastered items and that of partially master items were sufficient. Figures 2-5 (Pp. 11-14) depict the probability distributions of trichotomous data simulated from f(x) on page 4. Here, 10% refers to the percent of mastered items whereas 70% the partially mastered items. Then on page 15, Figure 6 combines various beta functions with 5 distinct test lengths. The overlay effect shows clearly that the shape of the compound beta-multinomial distribution is determined solely by parameters of the beta functions. The weight coefficient (w), as one might imagine, would not affect the probabilistic functions shown on pages 11-15.

When the percentages varied from a (10%, 70%) combination to a (30%, 30%) combination, the appearance of beta-multinomial distributions altered accordingly; although the general shape remained unchanged.

So, where is the beef? Sadly enough, the simulated data based on the beta-multinomial distributions did not depart significantly from those generated by the beta-binomial model (see Figures 9-13 for the univariate cases and Figure 14 for one bivariate case). Perhaps this was the main reason why Huynh preferred the beta-binomial model even for cases involving multiple classifications (e.g., Huynh, 1978, Psychometrika). His preference certainly should not preclude the utility of beta-multinomial models in the present context. Conceptually, the beta-multinomial model is well matched with the framework of a multiple classification, more so probably than the simple beta-binomial model. Before committing a fatal error in her conceptualization of the problem, the author welcomes insights or comments on her proposed methodology.

REFERENCES

Cheng, Ping. Minimax estimates of parameters of distributions belonging to the exponential family, Acta Mathematica Sinica, 1964, 5, 277-299.

Cohen, J.A. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20, 37-46.

Huynh, H. On the reliability of decision in domain-referenced testing. Journal of Educational Measurement, 1976, 13, 253-264.

IMSL Reference Manual (ed. 6). Houston, TX: International Mathematical and Statistical Librarys, Inc., 1977.

Ishii, G. and Hayakawa, R. On the compound binomial distribution, Annals of the Institute of Statistical Mathematics, Tokyo, 1960, 12, 69-80 (Errata, 12, 208).

Johnson, N.L. and Kotz, S. Distributions in Statistics: Discrete Distributions. Boston, MA: Houghton-Mifflin Company, 1969.

Mosimann, J.E. On the compound multinomial distribution, the multivariate β-distribution and correlations among proportions, Biometrika, 1962, 50, 47-54.

Peng, C.Y. and Subkoviak, M.J. An investigation of Huynh's normal approximation procedure for estimating criterion-referenced reliability. Journal of Educational Measurement, 1980, 17, 359-368. (Equal Authorship).

Subkoviak, M.J. Empirical investigation of procedures for estimating reliability for mastery tests. Journal of Educational Measurement, 1978, 15, 111-116.

Subkoviak, M.J. Decision-consistency approaches. In R.A. Berk (ed.), Criterion-Referenced Measurement: The state of the art. Baltimore, MD: Johns Hopkins Press, 1979.

UNIVARIATE PROBABILITY--BETATRINOMIAL

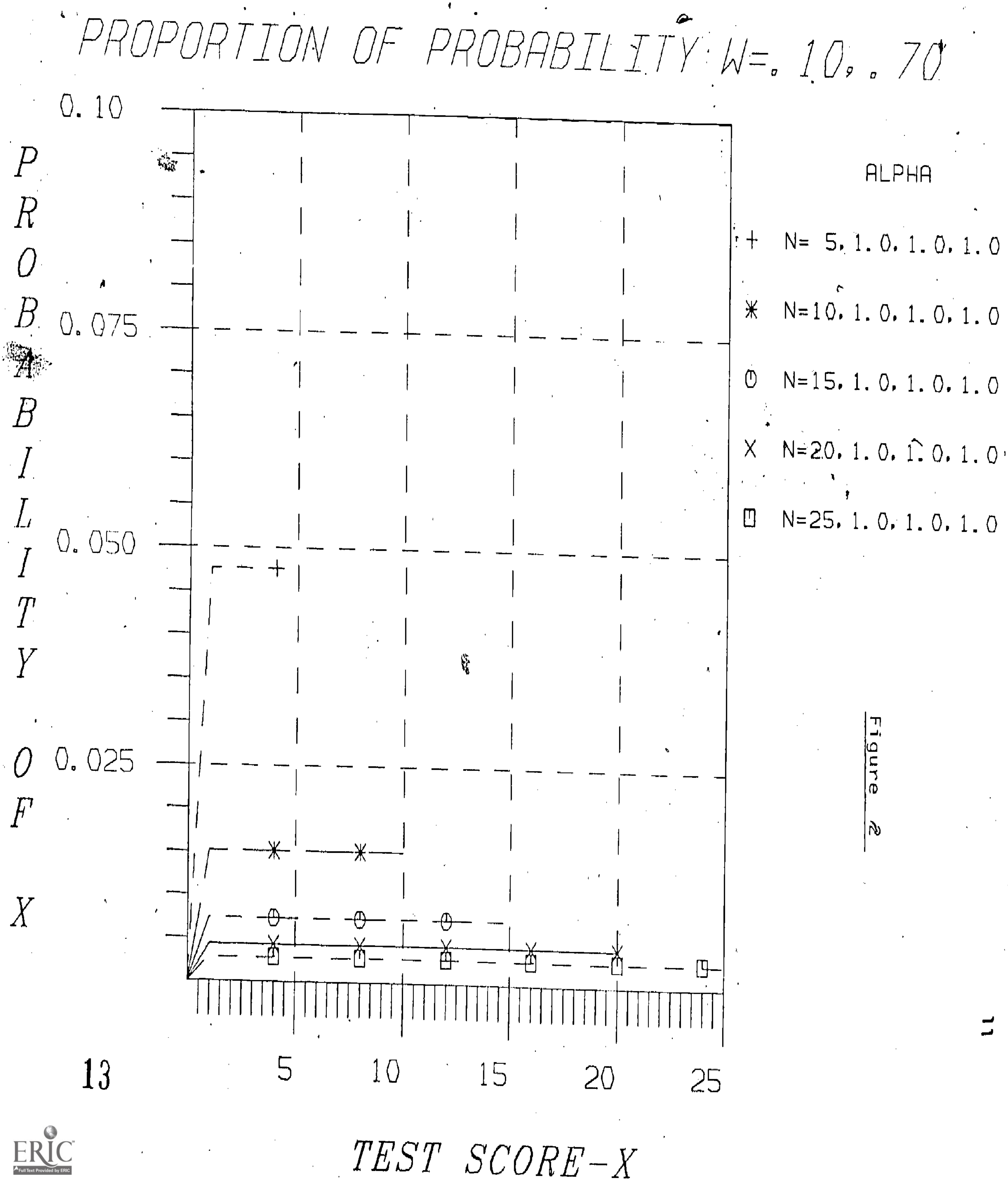


Figure 2

11

UNIVARIATE PROBABILITY--BETATRINOMIAL

PROPORTION OF PROBABILITY W=.10,.70

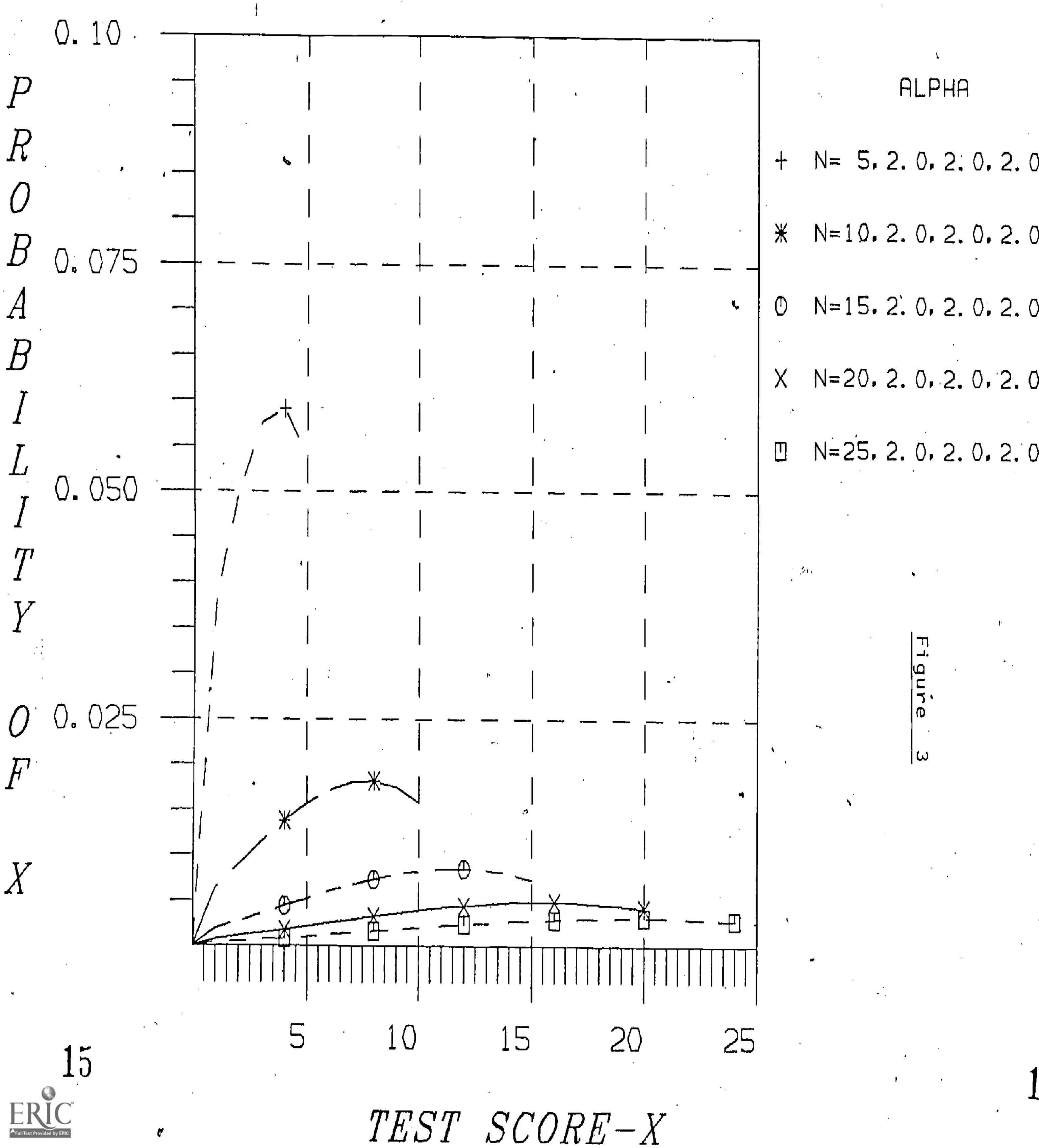


Figure 3

UNIVARIATE PROBABILITY--BETATRINOMIAL

PROPORTION OF PROBABILITY W=.10,.70

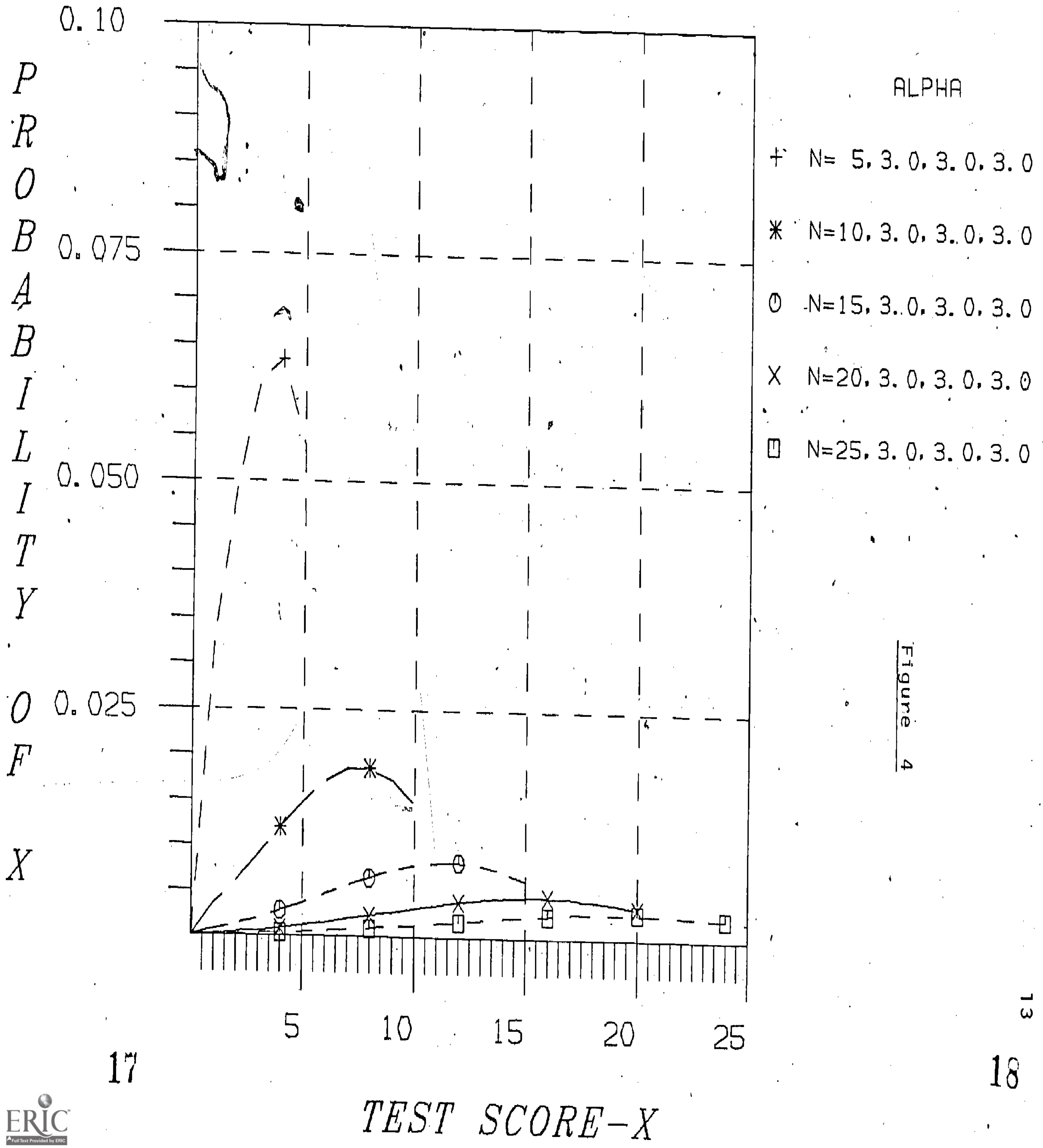


Figure 4

UNIVARIATE PROBABILITY--BETATRINOMIAL

PROPORTION OF PROBABILITY W=.10,.70

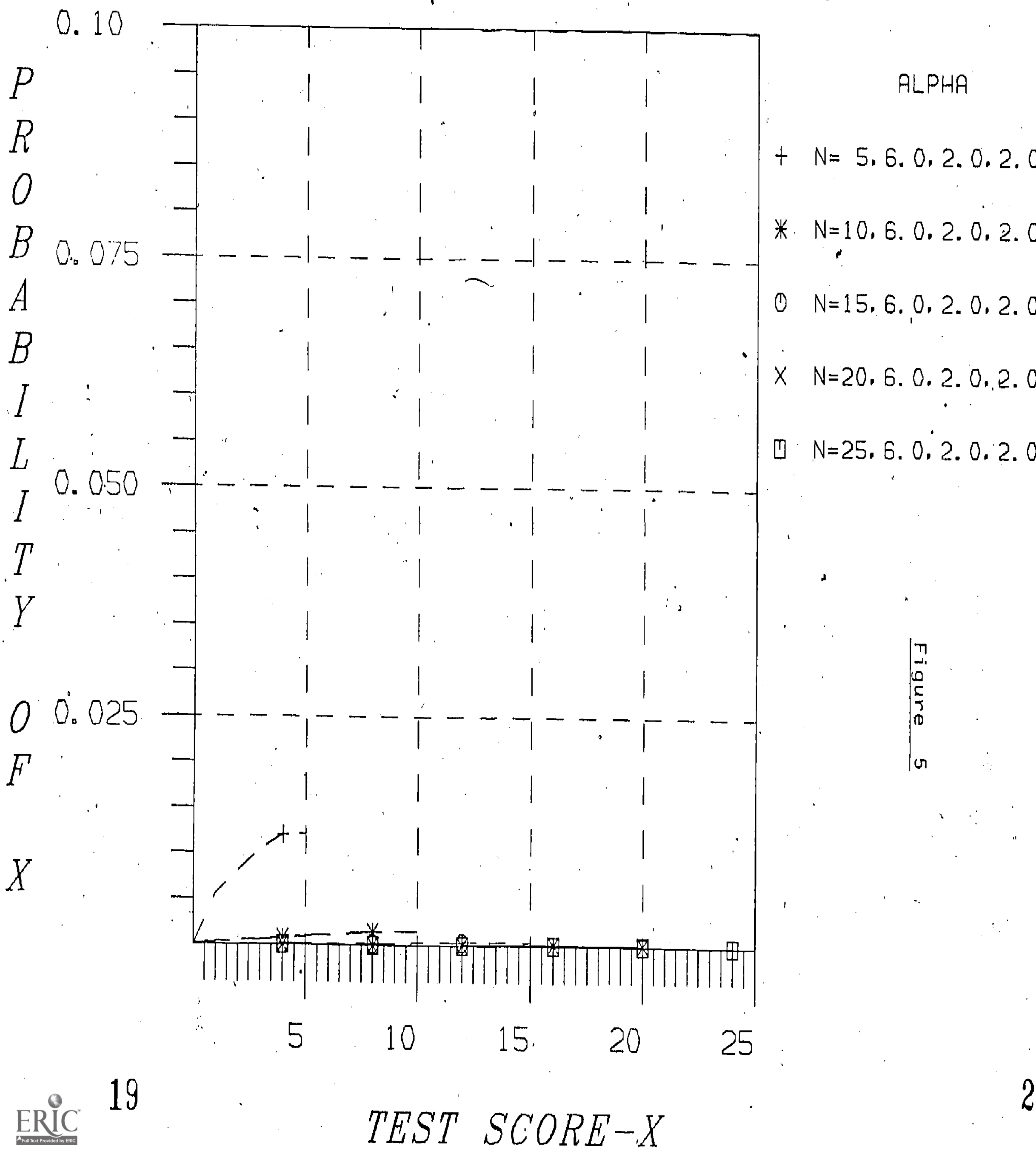


Figure 5

UNIVARIATE PROBABILITY--BETATRINOMIAL

PROPORTION OF PROBABILITY W=.10,.70

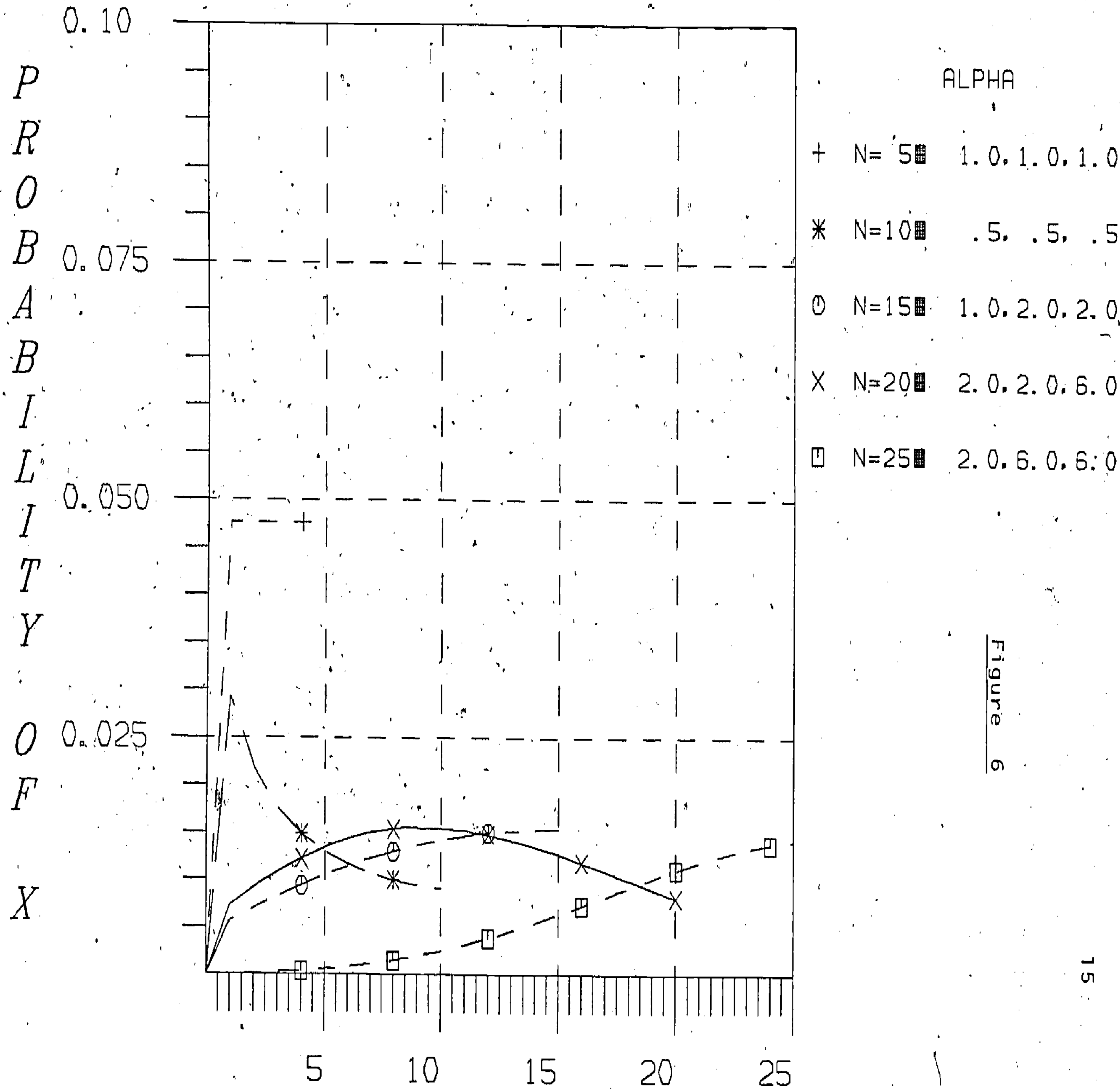
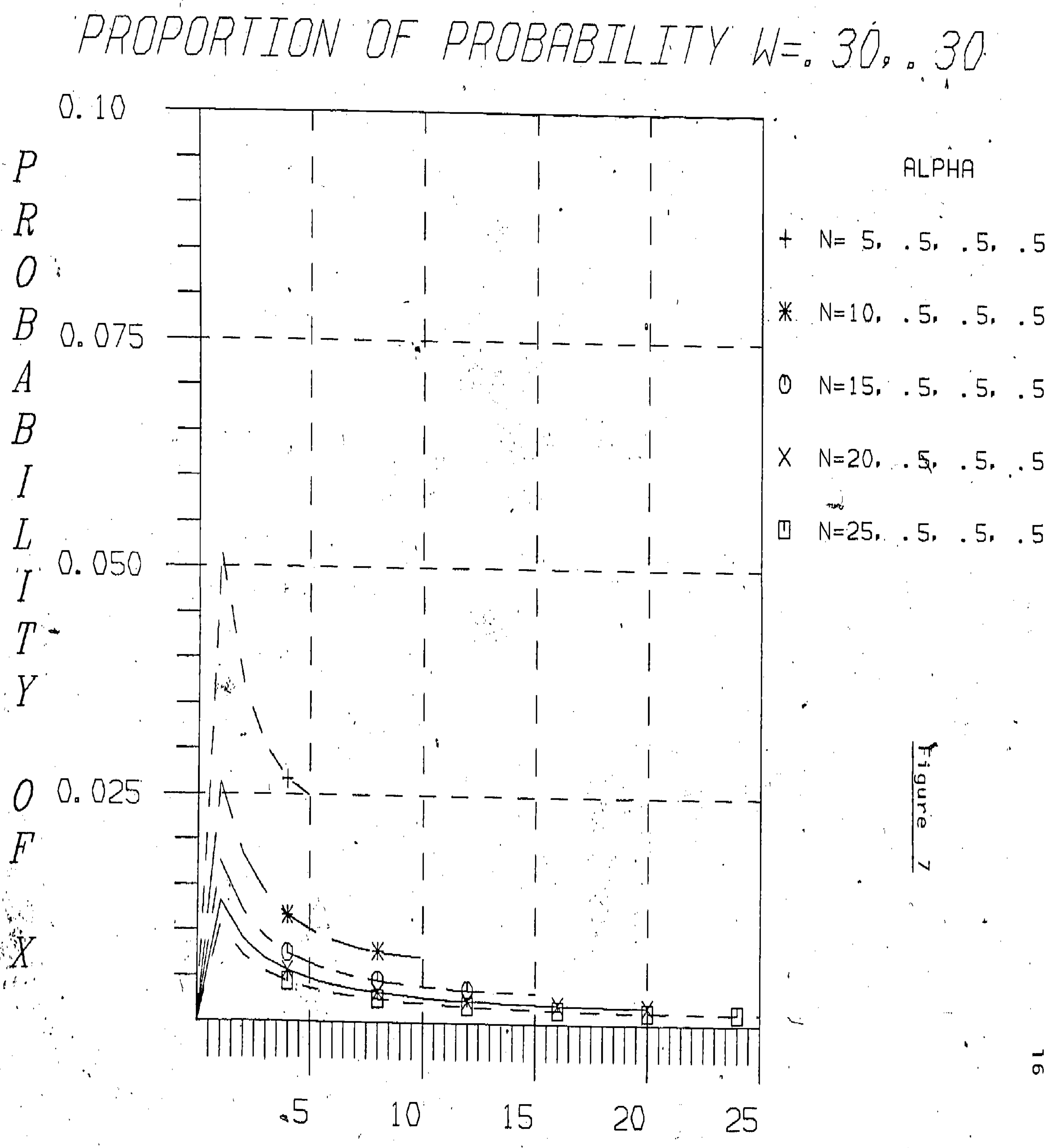


Figure 6

15

UNIVARIATE PROBABILITY--BETATRINOMIAL



TEST SCORE-X

Figure 7

UNIVARIATE PROBABILITY--BETATRINOMIAL

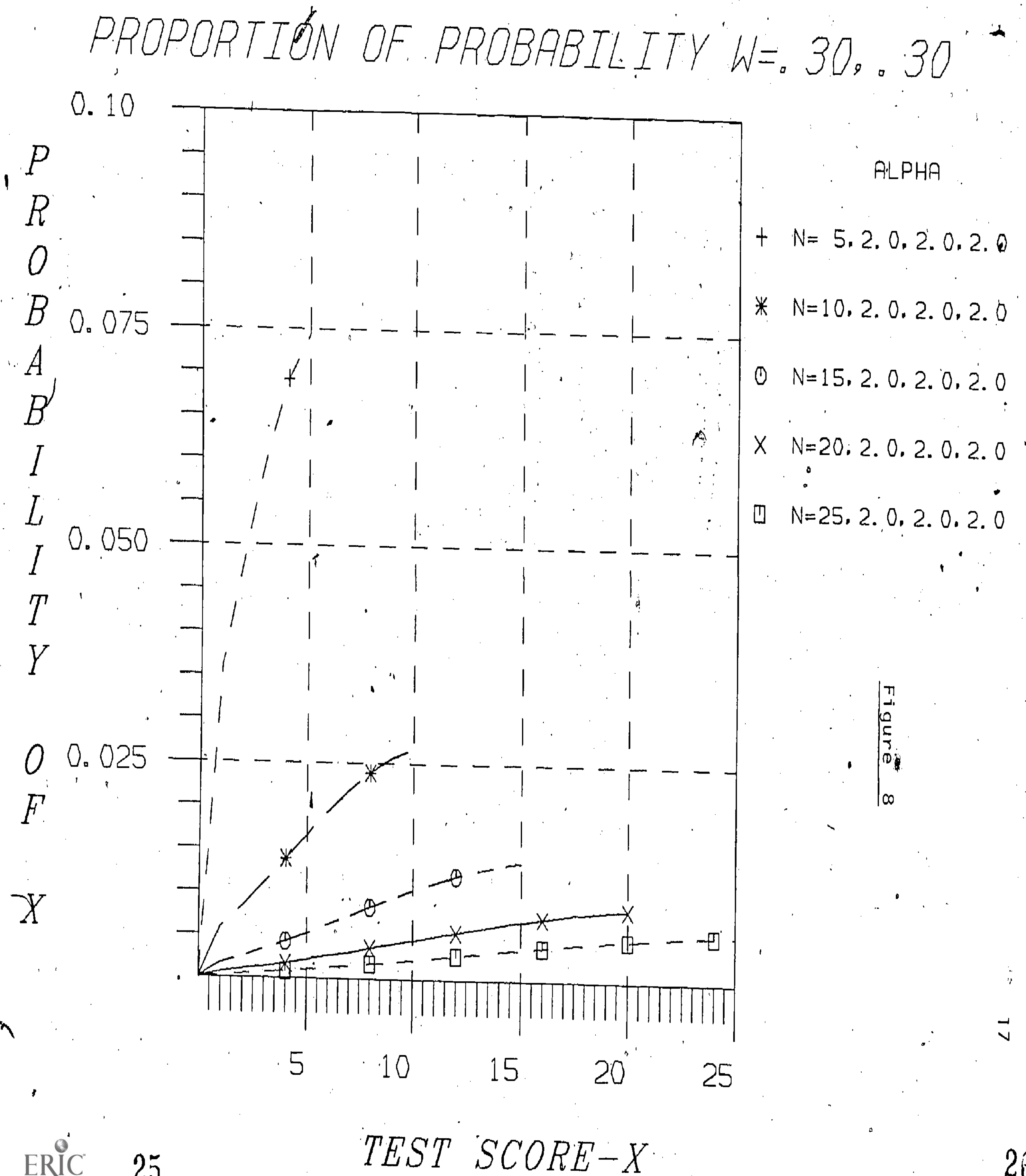


Figure 8

17

UNIVARIATE PROBABILITY--BETA + BINOMIAL
ALPHA=1, BETA=1, N=5, 10, 15, 20, 30

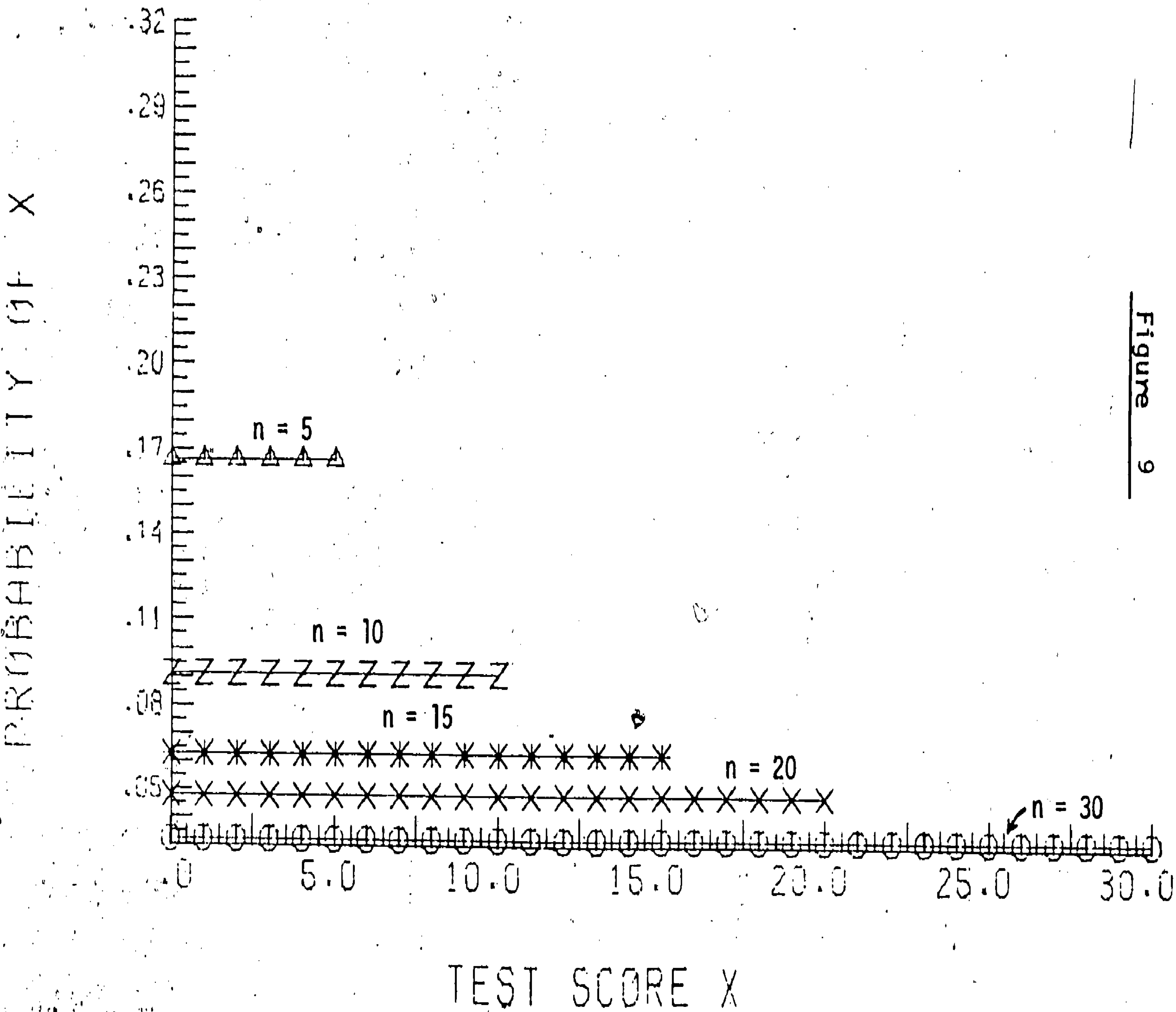


Figure 9

UNIVARIATE PROBABILITY--BETA + BINOMIAL
ALPHA=.5,BETA=.5,N=5,10,15,20,30

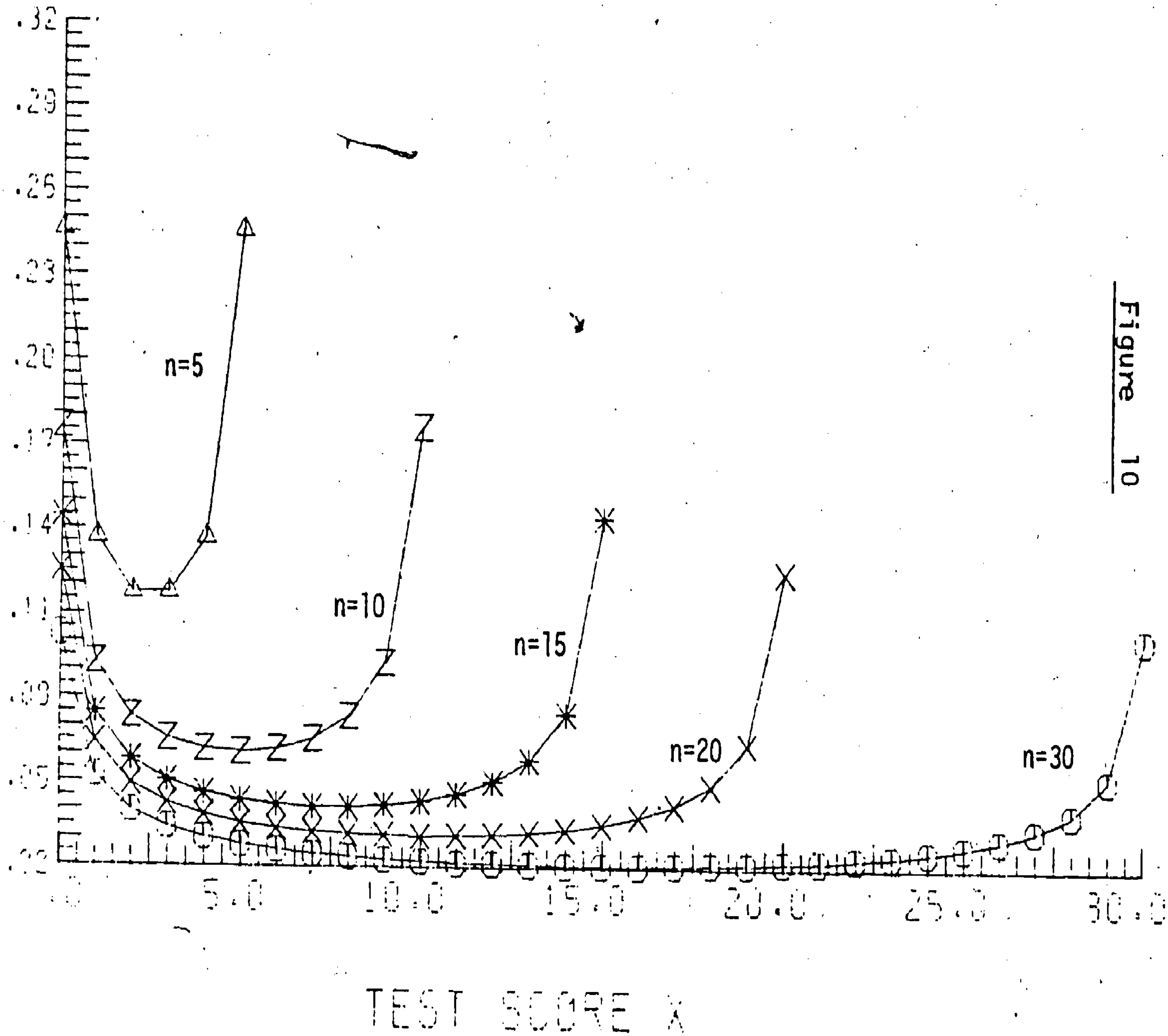


Figure 10

BEST COPY AVAILABLE

UNIVARIATE PROBABILITY--BETA + BINOMIAL
ALPHA=2, BETA=2, N=5, 10, 15, 20, 30

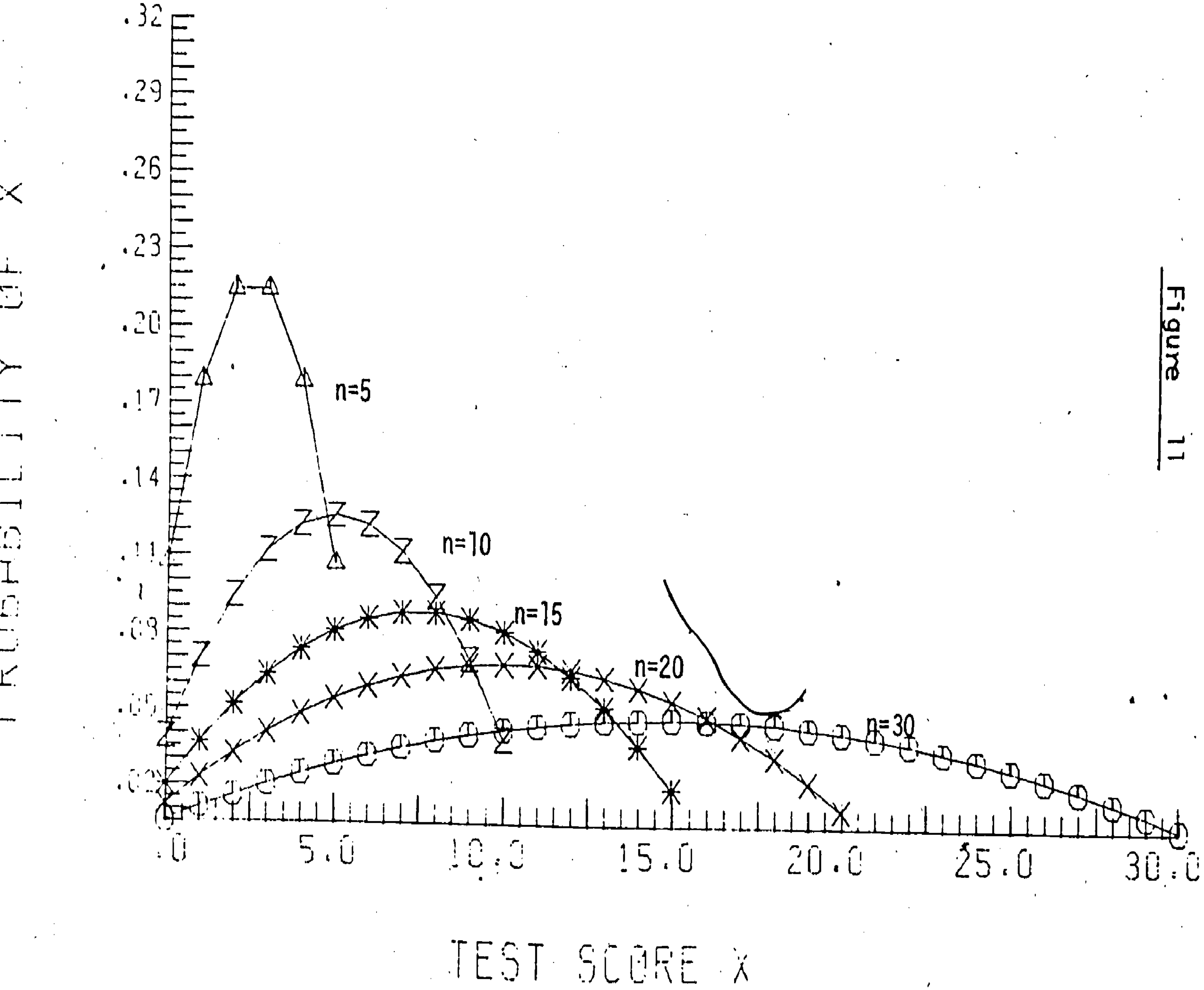


Figure 11

UNIVARIATE PROBABILITY--BETA + BINOMIAL
ALPHA=3, BETA=3, N=5, 10, 15, 20, 30

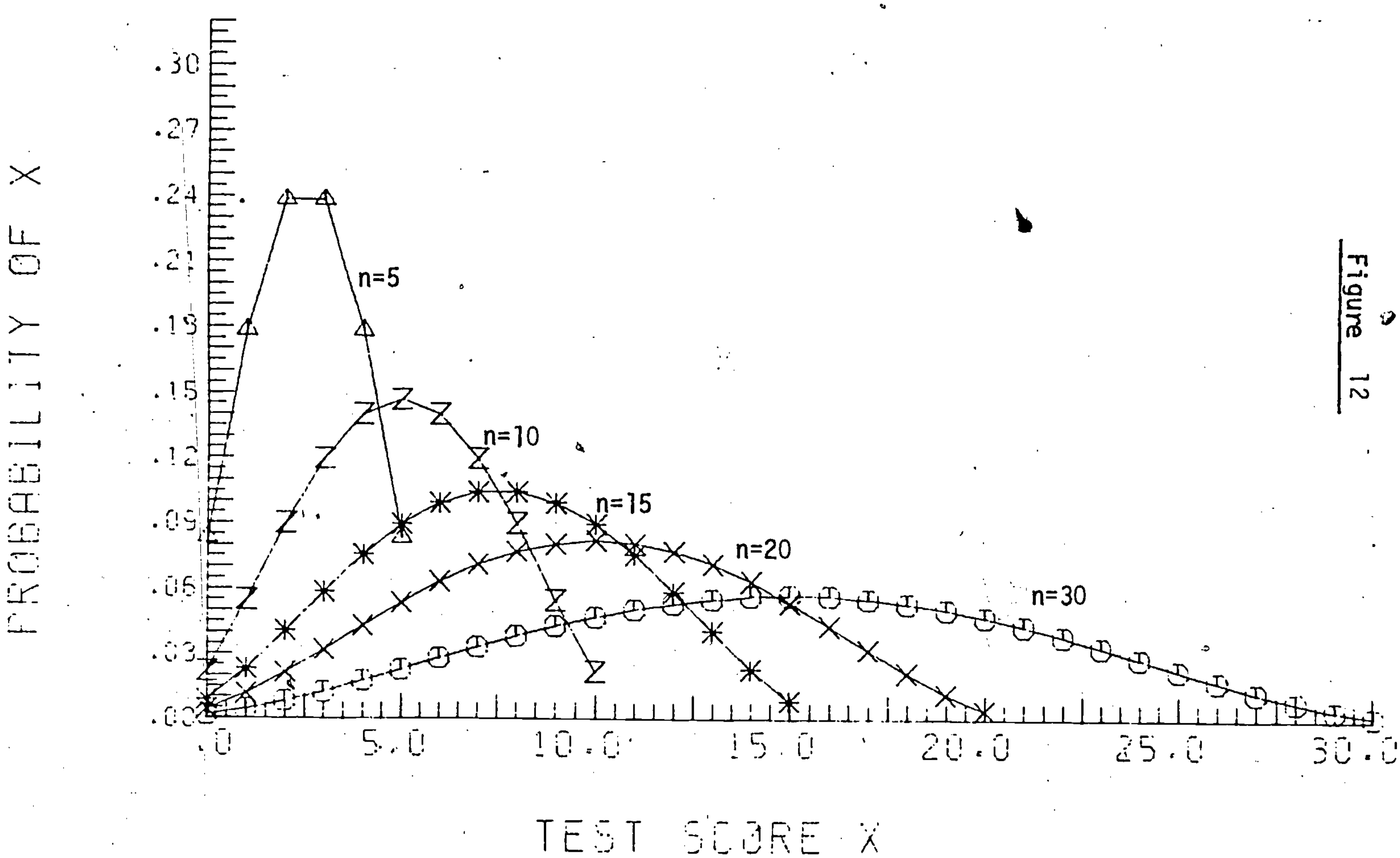


Figure 12

BEST COPY AVAILABLE

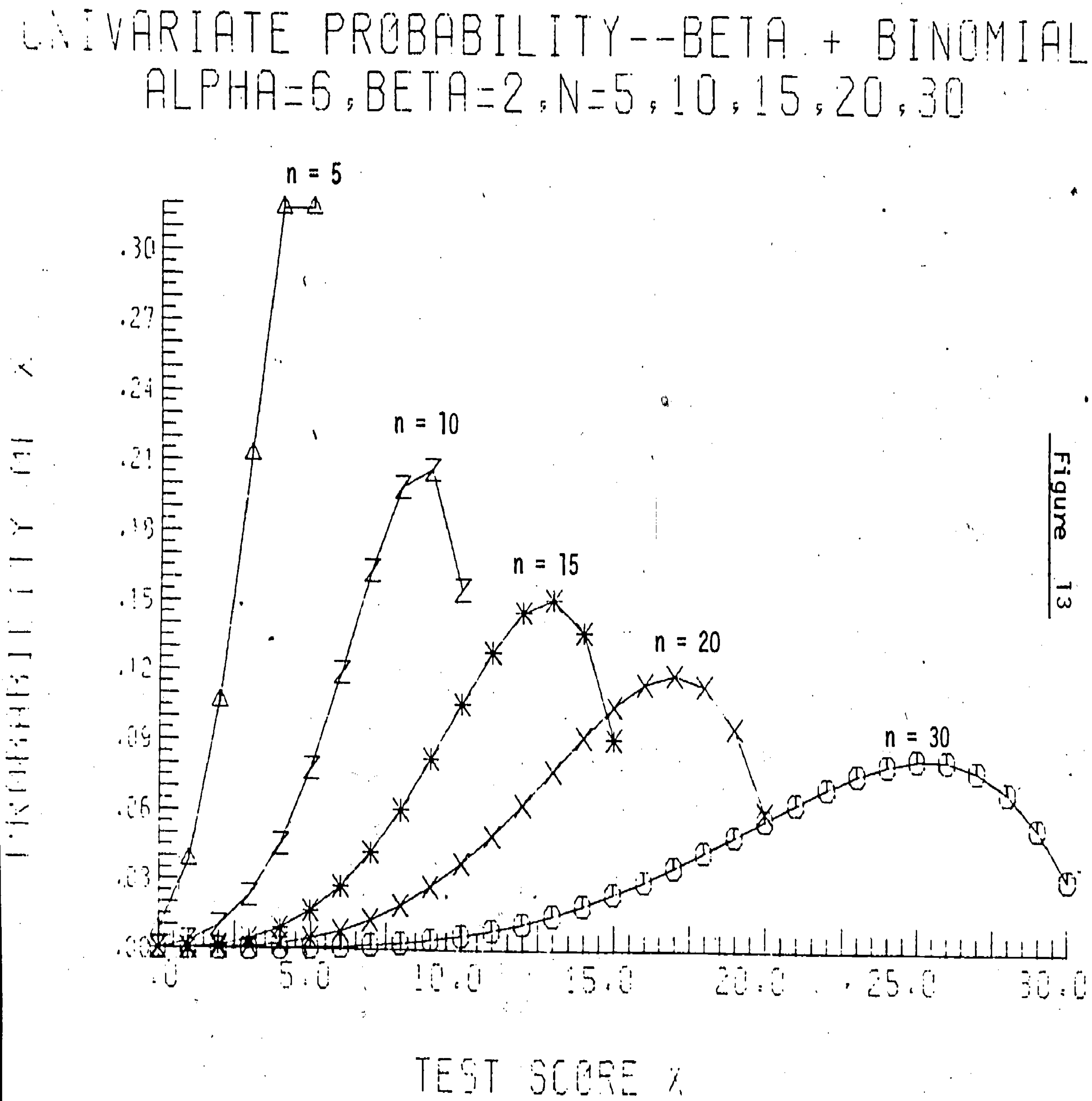
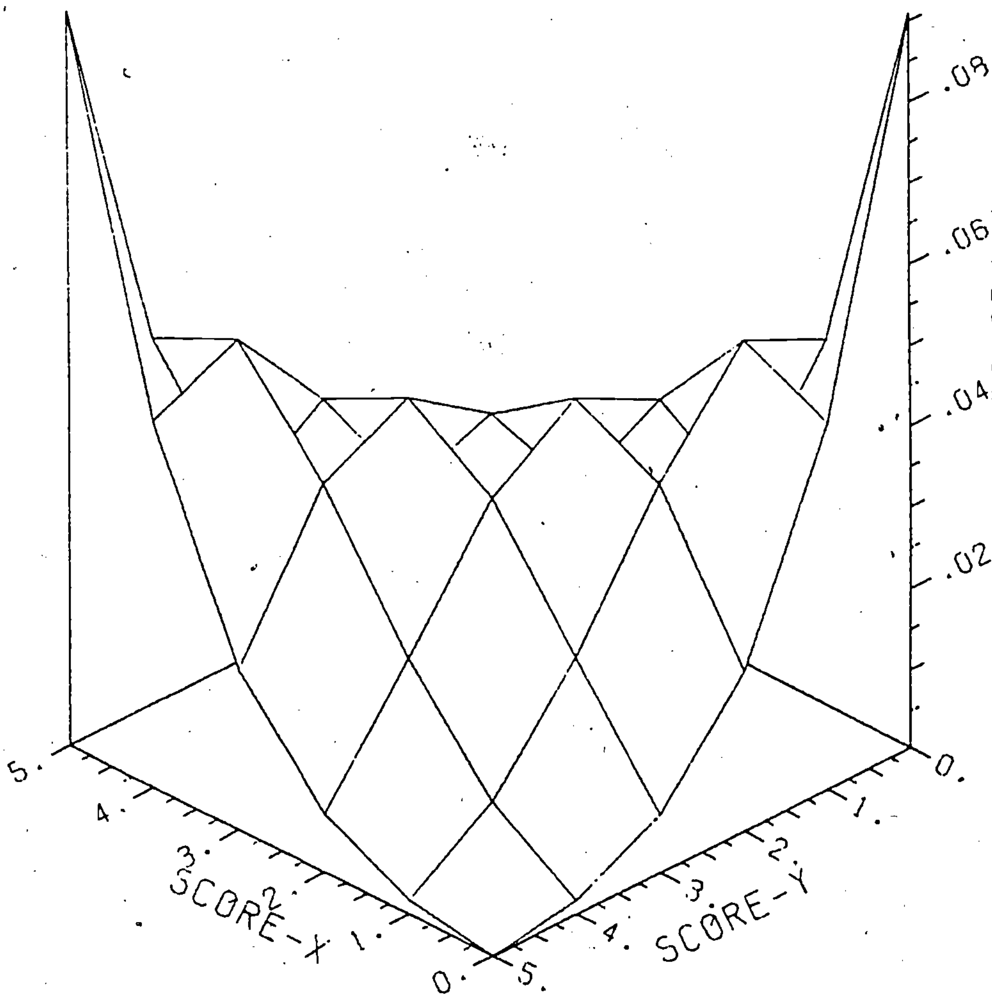


Figure 13

BEST COPY AVAILABLE

Figure 14



BIVARIATE BETA-BINOMIAL
ALPHA=1,BETA=1,N=5