DOCUMENT RESUME

ED 074 140 TM 002 503

AUTHOR Kleinke, David J.
TITLE Further Studies of Linear Prediction Following Matrix Sampling.
PUB DATE Feb 73
NOTE 8p.; Paper presented to joint session of American Educational Research Association and National Council on Measurement in Education (New Orleans, La., Feb. 25-March 1, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Item Sampling; *Norms; *Predictive Measurement; *Sampling; *Scores; Speeches; Technical Reports; True Scores

ABSTRACT

In a post mortem study, it is demonstrated that linear prediction is as effective as computing a negative hyper-geometric distribution for estimating test norms following matrix sampling from a total test with a highly skewed score distribution, provided the same prediction coefficient is used for all examinee groups. It is also demonstrated empirically and algebraically that using a coefficient unique to each examinee group produces distributions of predicted total-test scores with "insufficient" variance. Implications for measurement practice and statistical theory are discussed. (Authors)

FURTHER STUDIES OF LINEAR PREDICTION FOLLOWING MATRIX SAMPLING

David J. Kleinke

Syracuse University

Paper presented to joint session of American Educational Research Association and National Council on Measurement in Education; New Orleans, February 1973.

Abstract

In a post mortem study, it is demonstrated that linear prediction is as effective as computing a negative hypergeometric distribution for estimating test norms following matrix sampling from a total test with a highly skewed score distribution, provided the same prediction coefficient is used for all examinee groups. It is also revealed that using a coefficient unique to each examinee group produces distributions of predicted total-test scores with "insufficient" variance.

Kleinke (1972) has presented a technique for generating total-test score distributions following matrix sampling. After an examinee responds to a sample of the items, his score on the items with which he was not presented is predicted. In the present paper, results of two further investigations of the technique are presented.

Matrix sampling is that procedure under which examinees and items are simultaneously sampled. Each examinee, then, responds to only a sample of items from the total test. The advantage of this procedure for test norming, as first presented by Lord (1962), is that by shortening the amount of time each examinee must spend in the norming process, cooperation of the examinees and their supervisors is enhanced and the representativeness of the norms thus obtained is enhanced. A number of studies of matrix sampling have appeared, demonstrating that accurate estimates of the total-test mean and variance are possible. Of these studies, in only a few (e.g., Cooke & Stufflebeam, 1967; Kleinke, 1972; Lord, 1962) has the total-test score distribution been estimated. While the mean and the variance are obviously essential data, it is equally obvious that information about the overall distribution is at least helpful in test norming.

The technique suggested by Lord (1962) for generating a total-test distribution is to use the estimated mean and variance and the number of items in the total test to generate a negative hypergeometric distribution (H) for the scores. Kleinke's (1972) approach is based on total-test score estimates for the examinees. It uses a linear-prediction (LP) equation

to predict a score for each examinee on the composite of items with which he was not presented:

$$Y_i' = \left(\frac{\hat{s}_T}{s_X} - 1\right)(X_i - \overline{X}) + \hat{\overline{Y}} , \qquad (1)$$

where

- Y_i' is examinee i's score on all items not in subtest X, which is the subtest to which i responded,
- $\hat{s}_T$ is the estimated total-test standard deviation,
- s_X is the standard deviation of subtest X,
- X_i is i's score on X,
- $\overline{X}$ is the mean of X, and
- $\hat{\overline{Y}}$ is the estimated mean of composite Y.

An examinee's predicted total-test score, T_i' is merely the sum of X_i and Y_i' .

In Kleinke (1972), it was demonstrated that the distribution following LP was as adequate as that of a generated H, for the total test he used. That test had an underlying score distribution that was only slightly positively skewed, but was subsequently discovered to have an underlying true-score distribution that did not go to zero, because of the operation of chance success. For this reason, the H-generated distribution was systematically, but slightly, different from the criterion distribution.[1] At the same time, the LP distribution had small and non-systematic differences from the criterion distribution caused by the rounding of predicted scores to integers and the resulting phenomenon that, for many scores, no examinees were predicted.

Two questions raised in Kleinke (1972) were investigated in the resent study. First, could LP be used if the total-test score distribution were highly skewed? Second, what would be the effect of using $\hat{s}_{T(X)}$, the estimated total-test standard deviation for examinee-sample x, in place of

$\hat{s}_T$, the mean of these estimates, in Equation 1?

Skewed Distribution

Method

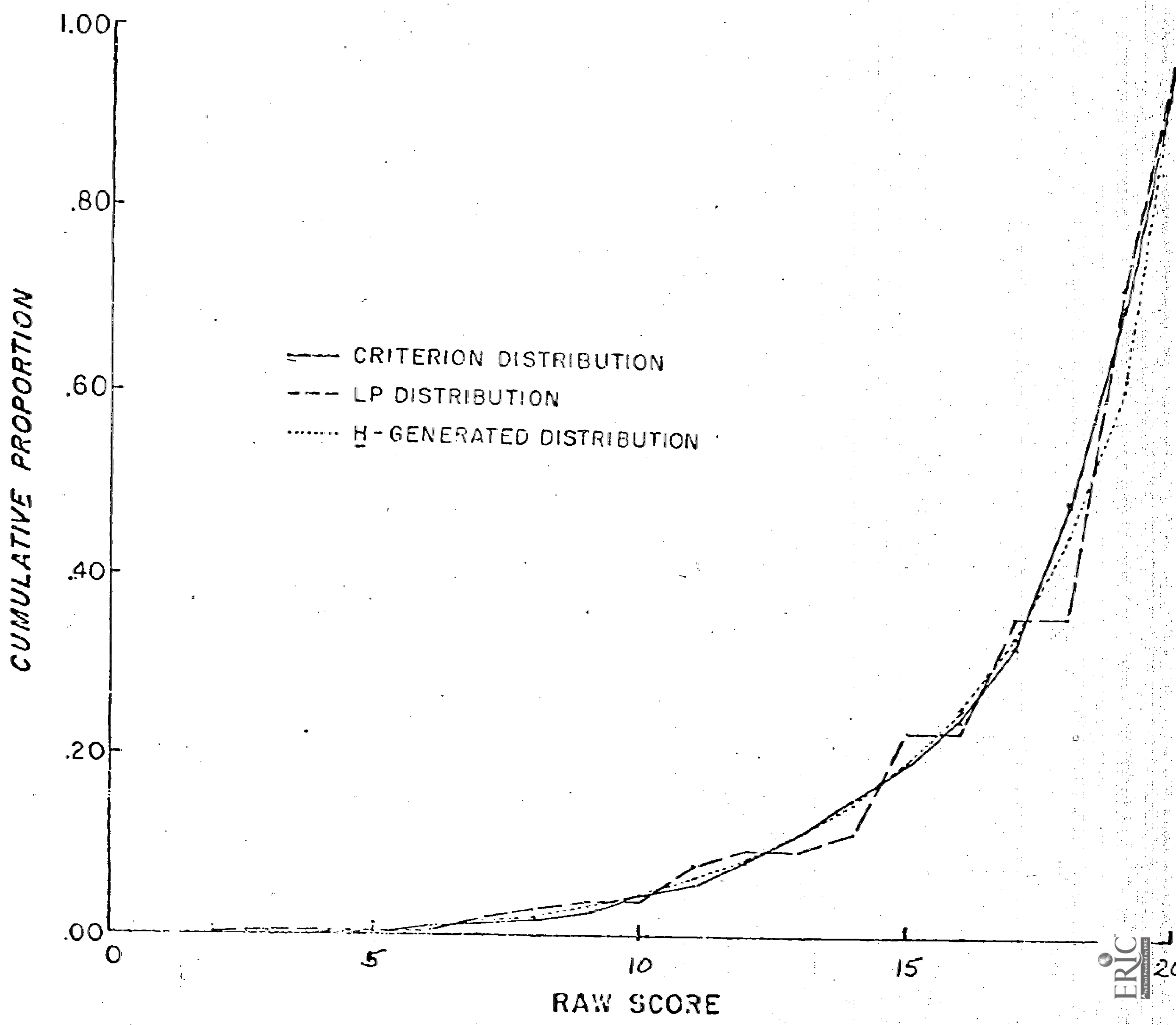
Item- and examinee-samples were drawn from the populations used by Shoemaker (1970).[2] As can be seen in Figure 1, the test was a twenty-item test with a J-shaped score distribution. Four item samples of five items each were selected randomly without replacement. The examinees were randomly assigned to one of four nonoverlapping examinee-groups, each containing 257 examinees. The procedures of Lord (1962) were used to obtain estimates of total-test mean and variance. Following this, both H and LP were used to approximate the total-test score distribution.

Results and Discussion

The actual mean and variance were 17.54 and 9.01, respectively. The corresponding matrix-sampling-based estimates were 17.60 and 10.07. Cumulative distribution curves are presented in Figure 1. Differences in proportions are presented in Table 1, together with the data from Kleinke's (1972) study, for comparative purposes.

Again, as in the original study, LP estimates suffer from the arbitrary rounding. Because of this rounding, no examinees were predicted to receive raw scores of 10, 13, 16, or 18. If even the first decimal were retained, the cumulative distribution would be much smoother and closer to the actual distribution. However, the original decision to concentrate on integer scores was based on two considerations. First, since the application of H provides for rounding estimated frequencies to the nearest examinee, it appeared reasonable to round LP-generated estimates to the nearest integer.

(4)



1.00
.80
.60
.40
.20
.00
CUMULATIVE PROPORTION
CRITERION DISTRIBUTION
LP DISTRIBUTION
H-GENERATED DISTRIBUTION
0
5
10
15
RAW SCORE

Table 1.

Maximum and mean differences between generated and criterion cumulative frequency distributions.

	Distribution			
	Skewed		Symmetric	
Difference	LP	H	LP	H
Maximum	.131	.076	.035	.041
Mean (Alg.)	-.001	.002	.003	.004
Mean (Abs.)	.021	.010	.010	.013

Given the present ratio of examinees to number of items in the total-test, this consideration seems to be inappropriately restrictive. However, the second consideration is still sufficiently strong to argue for continuation of the practice of rounding LP-scores. That is because the advantage of using LP over generating H lies in its potential for giving information to an examinee or to one who must make a decision on the basis of performance on only part of a set of scores, as in the case when a job applicant, for instance, has been presented with only a portion of a regular battery of tests, but the total score is used for decision-making. Expressing expected decimal scores on tests composed of binarily-scored items could lead to confusion and the appearance of greater precision than the data probably warrant.

In addition, it should be pointed out that the greatest discrepancies were observed close to the modal value. This was also true for the (nearly) symmetric distribution. On first consideration, this would seem to be a more serious problem than it probably is. In practice, the test constructor is advised to fit the difficulty level of a test and, hence, the shape of its score distribution, to the purpose for which the test is intended.

A highly skewed distribution, such as that of this test, would be appropriate where differences among low scorers were relatively more important than identifying differences among high scorers.

Use of $\hat{s}_{T(X)}$

Initially, it was suggested that using $\hat{s}_{T(X)}$ in place of $\hat{s}_T$ in Equation 1 would be more reasonable for predicting total-test scores from item-sample information. The rationale for that is that $\hat{s}_{T(X)}$ is based on the unique information that arises from the responses of members of examinee-group X and the particular set of items with which they were presented.

However, when $\hat{s}_{T(X)}$ was used, the variances of the predicted total-test scores were all less than $\hat{s}_T$. Some reflection and numerical and algebraic manipulation revealed why. Using $\hat{s}_{T(X)}$ is essentially equivalent to averaging (and hence, summing) essentially non-additive standard deviations.

Footnotes

1 The author is indebted to Frederic M. Lord who pointed out the true-score distribution and its effect on H, after examining the data.

2 The author wishes to express his thanks to David M. Shoemaker for providing him with data.

REFERENCES

Cook, D. L. and Stufflebeam, D. L. Estimating test norms from variable size item and examinee samples. Journal of Educational Measurement, 1967, 4, 27-33.

Kleinke, D. J. A linear-prediction approach to developing test norms based on matrix-sampling. Educational and Psychological Measurement, 1972, 32, 75-84.

Lord, F. M. Estimating norms by item-sampling. Educational and Psychological Measurement, 1962, 22, 259-267.

Shoemaker, D. M. Item-examinee sampling procedures and associated standard errors in estimating test parameters. Journal of Educational Measurement, 1970, 7, 255-262.