PSYCHOMETRIKA—VOL. 27, NO. 1
MARCH, 1962

A THEORETICAL DISTRIBUTION FOR MENTAL TEST SCORES

J. A. KEATS
UNIVERSITY OF QUEENSLAND, AUSTRALIA

AND

FREDERIC M. LORD
EDUCATIONAL TESTING SERVICE*

The negative hypergeometric distribution of raw scores on mental tests is derived from certain assumptions relating to test theory. This result is checked empirically in a number of examples. Further derivations lead to the bivariate distribution of parallel tests which is also verified with actual data. The bivariate distribution of raw score and true score is also derived from a further assumption. This distribution is used to set confidence limits for true scores for persons with a given raw score.

In an earlier publication Keats [1] expressed the view that mental test scores could be adequately represented by the hypergeometric distribution with a negative parameter. This distribution was referred to as the *hypogeometric* distribution, but will be referred to here as the *negative hypergeometric* distribution. In that account the reasons for choosing this particular distribution were rather intuitive. Attempts to investigate this distribution produced many interesting results relating to test theory, but failed to show a logical link between the usual assumptions of test theory and the distribution which, from empirical study, was known to provide a good model for scaling tests—in the sense of providing accurate centile scores and age corrections. The present article will relate the derivation of the hypergeometric distribution to certain test theory assumptions. Further properties of this distribution as well as certain bivariate distributions are derived and illustrated with actual data. Let

x = raw score,
p = relative true score,
n = number of items in the test,
N = number of subjects in the sample,
u = an arbitrary variable used to define the factorial moment generating function $M[x, u]$,

*This work was supported in part by contract Nonr-2752(00) between the Office of Naval Research and Educational Testing Service. Reproduction in whole or in part for any purpose of the United States Government is permitted.

$P(x, p)$ = the probability of x and p,
b = the slope of the regression line of p on x,
r = Kuder-Richardson reliability formula 21,
$g(x)$ = frequency distribution of x,
$f(p)$ = frequency distribution of p.

Assume that at any given true-score level, p, the errors of measurement are independent and have a binomial distribution with parameter p. If the distribution of true scores in a given group of examinees is $f(p)$, then the distribution of raw scores will be

$$g(x) = \int_0^1 \binom{n}{x} p^x q^{n-x} f(p)\, dp,$$

where $q = (1 - p)$.

The equation just given will hold exactly in the special case where (i) all items have one and only one common factor, and (ii) for people at any given ability level, all items are of equal difficulty (i.e., all items have the same item characteristic curve). The equation should provide a good approximation in cases where neither (i) nor (ii) hold, as discussed in [3] and [4].* In the final analysis, the value of the equation will, of course, be judged by the accuracy with which conclusions drawn from it are found to fit a variety of real data.

The factorial moment generating function of x is

$$\begin{aligned} M[x, u] &= \sum_0^n (1 + u)^x \int_0^1 \binom{n}{x} p^x q^{n-x} f(p)\, dp \\ &= \int_0^1 \sum_0^n \binom{n}{x} (1 + u)^x p^x q^{n-x} f(p)\, dp \\ &= \int_0^1 (1 + up)^n f(p)\, dp \\ &= \int_0^1 f(p)\, dp + nu \int_0^1 pf(p)\, dp + \cdots \\ &\quad + \frac{n!}{(n - r)!} \frac{u^r}{r!} \int_0^1 p^r f(p)\, dp + \cdots + u^n \int_0^1 p^n f(p)\, dp. \end{aligned}$$

The right-hand side of this expression has for the coefficients of $u^r/r!$ the moments about the origin of the distribution of p multiplied by $n!/(n-r)!$;

*This approximation is believed to be a good one for the present purpose of finding practical mathematical forms for representing the distribution of x and the distribution of p. The fact that this approximation leads in several cases to the Kuder-Richardson formula-21 reliability coefficient, as will be shown in due course, is not to be interpreted as suggesting the use of this formula in place of formula 20, however. Such a choice of reliability formulas should be based on test reliability theory, which is a much more fully developed theory than any presently available for representing $g(x)$ and $f(p)$.

but these coefficients in the expansion of $M[x, u]$ are the factorial moments of $g(x)$. Hence the moments about the origin of $f(p)$ can be written down in terms of the factorial moments of $g(x)$ and so of the ordinary moments of $g(x)$ together with a function of the constant n. The relationships are

$$M_1'(p) = \frac{M_1'(x)}{n},$$

$$M_2'(p) = \frac{M_2'(x) - M_1'(x)}{n(n-1)},$$

$$M_3'(p) = \frac{M_3'(x) - 3M_2'(x) + 2M_1'(x)}{n(n-1)(n-2)},$$

$$M_4'(p) = \frac{M_4'(x) - 6M_3'(x) + 11M_2'(x) - 6M_1'(x)}{n(n-1)(n-2)(n-3)},$$

and so on.

In particular,

$$\text{Mean }(p) = \frac{\text{Mean }(x)}{n};$$

$$\text{Variance }(p) = \frac{1}{n(n-1)}\,[\text{Variance }(x)] - \frac{\text{Mean }(x)}{n}\,[n - \text{Mean }(x)].$$

This last result has some theoretical interest. Let $y = np$, so that y is the number-correct true score. Then

$$\frac{\text{Var }(y)}{\text{Var }(x)} = \frac{n}{n-1}\,\frac{\text{Var }(x) - n\bar{p}\bar{q}}{\text{Var }(x)}.$$

The right-hand side of this formula is thus the Kuder-Richardson formula 21 for reliability and can be interpreted as the ratio of the variance of the underlying probability distribution with range 0 to n to the variance of obtained scores.

In a similar way the first product moment and hence the correlation between p and x can be found. The result obtained is the square root of the Kuder-Richardson formula 21 as expected.

The derivation of the negative hypergeometric distribution will now be given. The original assumptions imply that $P(x, p)$, the joint distribution of p and x, is

$$P(x, p) = \binom{n}{x} p^x q^{n-x} f(p).$$

Consequently, the conditional distribution of p for given x is

$$P(p \mid x) = \frac{\binom{n}{x} p^x q^{n-x} f(p)}{g(x)},$$

where $g(x)$ is the frequency distribution of x. In particular

$$\text{Mean } (p \mid x) = \frac{1}{g(x)} \int_0^1 \binom{n}{x} p^{x+1} q^{n-x} f(p)\, dp.$$

A little algebraic manipulation shows that $1 - \text{Mean } (p|x) = \text{Mean } (1 - p|x)$, i.e.,

$$1 - \text{Mean } (p \mid x) = \frac{1}{g(x)} \int_0^1 \binom{n}{x} p^{x} q^{n-x+1} f(p)\, dp.$$

Thus,

$$g(x)[1 - \text{Mean } (p \mid x)] = \frac{n - x + 1}{x} \text{Mean } (p \mid x - 1) g(x - 1).$$

This is an important relationship between the frequency distribution of raw scores and the regression (linear or curvilinear) of p on x. If the regression is assumed to be *linear*, then $g(x)$ can be explicitly stated. If the regression of p on x is linear, then

$$\text{Mean } (p \mid x) = \text{Mean } (p) + b(x - \bar{x}),$$

where b is the usual regression coefficient of p on x; but

$$\text{Mean } (p) = \frac{\bar{x}}{n} \quad \text{and} \quad b = \frac{r}{n},$$

when r is the Kuder-Richardson formula 21. Hence

$$g(x)\left(1 - \frac{(1 - r)\bar{x}}{n} - \frac{rx}{n}\right) = \frac{n - x + 1}{x}\left[\frac{\bar{x}}{n} + \frac{r}{n}(x - 1 - \bar{x})\right] g(x - 1),$$

$$g(x) = \frac{(n - x + 1)(rx + (1 - r)\bar{x} - r)}{x(n - (1 - r)\bar{x} - rx)} g(x - 1)$$

$$= \frac{[n - (x - 1)]}{x} \cdot \frac{M\bar{x} + n(x - 1)}{M(n - \bar{x}) + n(n - x)} \cdot g(x - 1),$$

where

$$r = \frac{n}{n + M} \quad \text{or} \quad M = \frac{n(1 - r)}{r}.$$

Then it may be shown that

$$(1) \qquad g(x) = K\binom{n}{x}[(M\bar{x})(M\bar{x} + n) \cdots (M\bar{x} + nx - n)][\{M(n - \bar{x})\}$$
$$\{M(n - \bar{x}) + n\} \cdots \{M(n - \bar{x}) + n(n - x - 1)\}],$$

where

$$\frac{1}{K} = M(M + 1) \cdots (M + n - 1)n^n.$$

This distribution is expressed in terms of constants all of which are either known or can be estimated from the data; $\bar{x}$ represents the population mean of x which is estimated by the sample mean. The distribution is hypergeometric in form with a negative parameter; it will be denoted here by $H(x)$,

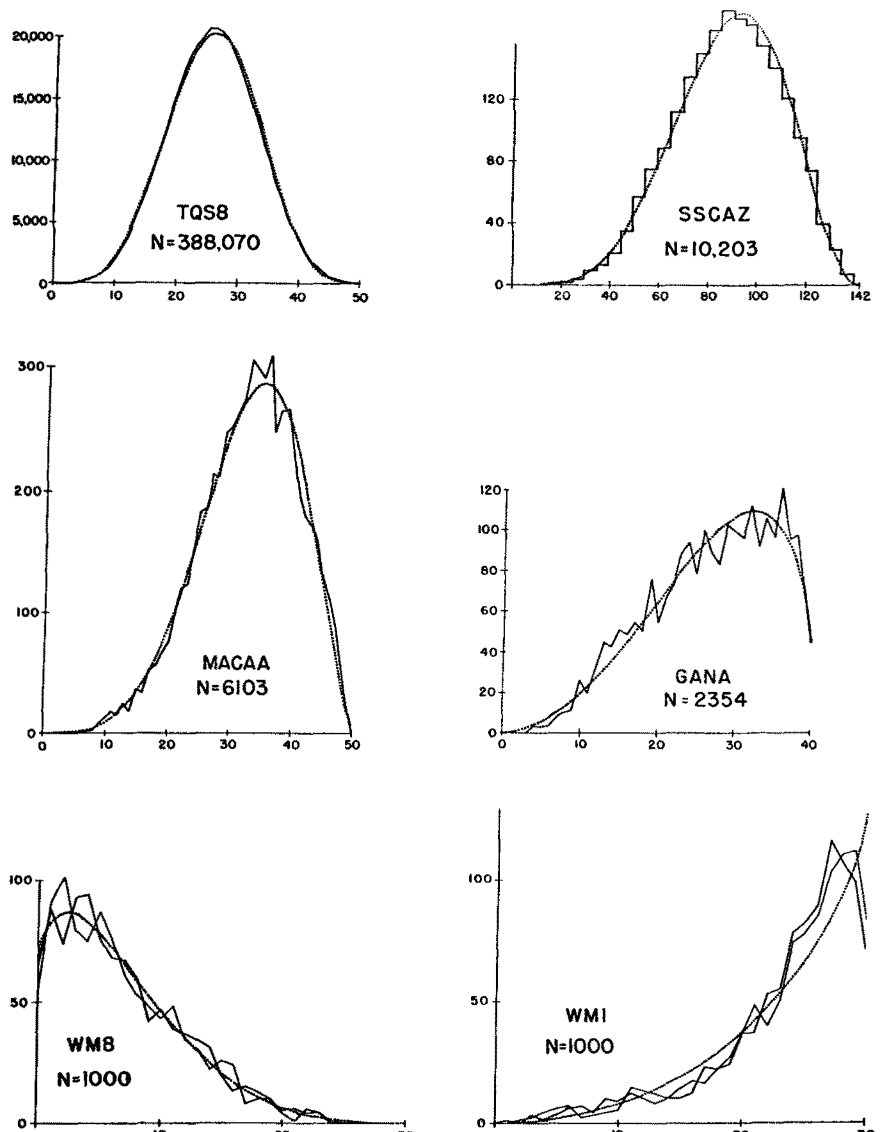

FIGURE 1
Negative Hypergeometric Distributions (dotted lines) Fitted to Six Sets of Test-Score Data

or simply by H. (A discussion of some of the properties of this distribution and of a convenient method for computing its frequencies will be found in [8, sec. 7.11] where it is called the beta-binomial distribution.) Thus on quite general assumptions this distribution could be expected to give a reasonable representation of actual test score distributions. The fact that it does in many cases is reported in [1]. Further verification is presented below.

A variety of test-score distributions were selected from nation-wide psychological testing programmes to represent a wide range of differently skewed shapes; a negative hypergeometric distribution was fitted to each. The first four distributions in Fig. 1 cover the full range of shapes and display as bad fits as any obtained for these data. The last two sets of data in Fig. 1 were selected from some experimental tests that Mollenkopf [6] had specially constructed to produce peculiarly shaped distributions. (Certain other distributions developed by Mollenkopf to be leptokurtic and symmetric ([6], p. 211) are not considered here, since it is known in advance that a symmetric negative hypergeometric distribution, $H(x)$, is always platykurtic—and, for that matter, so are symmetric test-score distributions in actual practice [5].)

In the diagrams, the solid-line frequency polygons represent ungrouped observed distributions, the histogram represents grouped observed data. The fitted hypergeometric distributions are represented by dotted "frequency polygons" in which the angles have been rounded off slightly so that they appear to be smooth curves. Frequency polygons are used for graphic clarity; the reader must keep in mind that all these distributions really represent discrete, not continuous variables.

Relevant statistics for the six sets of data are summarized in Table 1.

TABLE 1

Statistics for Six Distributions

Test code	Number of cases (N)	Mean ($\bar{x}$)	Standard deviation (s_x)	r	Parameters for H: n	$\frac{M\bar{x}}{n}$	$\frac{M(n-\bar{x})}{n}$	Chi square (χ^2)	Significance** level for χ^2
TQS8	388,071	25.82	7.28	.780	50	7.269	6.81	*	$.001 > P(\chi^2)$
SSCAZ	10,203	88.08	21.73	.936	142	6.047	3.70	31.7	$.10 > P(\chi^2_{21}) > .05$
MACAA	6,103	32.93	8.04	.843	50	6.135	3.18	50.9	$.10 > P(\chi^2_{38}) > .05$
GANA	2,354	27.06	8.19	.892	40	3.283	1.57	50.6	$P(\chi^2_{38}) \doteq .025$
WM8	1,000	6.76 6.75	5.12 5.11	.830	30	1.386	4.77	24.1 20.2	$.70 > P(\chi^2_{28}) > .50$ $.90 > P(\chi^2_{28}) > .80$
WM1	1,000	23.75 23.88	5.59 5.62	.873	30	3.457	0.90	84.9 *	$.001 > P(\chi^2_{19})$ *

*Not computed.

**These levels underestimate the obtainable fit (see text).

For each of the last two sets of data, there are two very similar empirical frequency distributions; for each set, a single H has been fitted, using appropriate averages of the empirical means and of the empirical standard deviations.

The discrepancies between the first set of data (TQS8) and the fitted H can be seen without much computation to be highly significant, since there are more than 20,000 cases for each value of x in the neighbourhood of the mode (the best fitting normal curve, incidentally, gives a slightly worse fit, lying somewhat further above the mode of the TQS8 distribution than H lies below). The last set of data (WM1) also gives a highly significant chi square, presumably because its shape is so extreme. Three of the remaining chi squares are fairly close to the five percent significance level. However, it should be remembered that each distribution has 1,000 or more cases. Moreover, the chi-square significance test underestimates the fit obtainable for $H(x)$, because the method used here (the method of moments) for estimating the parameters of H is not fully efficient, especially for highly skewed distributions. Examination of Fig. 1 suggests that all these fits may be adequate for many purposes. In fact, when it is considered that *no measure of the skewness of the observed data has been used in the process of fitting H* (the parameters of H were computed from the mean, the standard deviation, and the upper bound of the observed variable), the fits obtained appear to be surprisingly good for such a wide variety of distributional shapes.

Under the present assumptions the Kuder-Richardson coefficient becomes

$$r = \frac{n}{M + n},$$

and this may be tested for significance from zero using chi square in the following way:

$$\chi^2_{N-1} = \frac{(N - 1)(M + n)}{M + 1},$$

where N is the number of persons in the sample [1].

It can be shown that the Spearman-Brown correction formula applies to the Kuder-Richardson formula under the present assumptions. Suppose that a test is increased in length from n_1 to n_2 items by the addition of items of the same type as were in the original test. Then it is reasonable to assume that M is the same for the new test as it was for the old.

Then

$$r_1 = \frac{n_1}{M + n_1} \quad \text{and} \quad r_2 = \frac{n_2}{M + n_2},$$

and after equating values of M and simplifying

$$r_2 = \frac{kr_1}{(k - 1)r_1 + 1} \quad \text{where} \quad k = \frac{n_2}{n_1}.$$

Since it is assumed that a person's score is determined by a probability figure related to his ability plus random fluctuation, it follows that when p is fixed, performance on one item is independent of performance on any other item. It is thus possible to divide a test into two parallel parts and observe the relationship between the scores obtained for each part.

Let x_0 and x_1 represent the scores on the two parts of the test which have n_0 and n_1 items respectively, i.e., $x_0 + x_1 = x$ and $n_0 + n_1 = n$.

It is assumed that

$$P(x_0 \mid p) = \binom{n_0}{x_0} p^{x_0} q^{n_0 - x_0},$$

$$P(x_1 \mid p) = \binom{n_1}{x_1} p^{x_1} q^{n_1 - x_1},$$

and since

$$P(x_0 , x_1 , p) = P(x_0 \mid p) P(x_1 \mid p) f(p),$$

it is possible to determine the bivariate distribution $P(x_0 , x_1)$ by integrating the expression for $P(x_0 , x_1 , p)$ with respect to p over the range (0, 1). When this is done and the resulting expression is simplified,

$$(2) \qquad P(x_0 , x_1) = k\binom{n_0}{x_0}\binom{n_1}{x_1}[\{M\bar{x}\}\{M\bar{x} + n\} \cdots \{M\bar{x} + n(x_0 + x_1 - 1)\}] \cdot [\{M(n - \bar{x})\}\{M(n - \bar{x}) + n\} \cdots \cdot \{M(n - \bar{x}) + n(n_0 + n_1 - x_0 - x_1 - 1)\}],$$

where

$$\frac{1}{k} = M(M + 1) \cdots (M + n_0 + n_1 - 1)n^n.$$

This result gives the theoretical bivariate distribution of x_0 and x_1 which may be compared with the bivariate distribution actually obtained by splitting the test into two parts. (This comparison will be made for one set of empirical data in the second paragraph below.) The parameters of this distribution are those already estimated for the univariate distribution together with the lengths of the two parts of the original test.

The conditional distribution of x_1 on x_0 may be derived from the equations already presented:

$$P(x_1 \mid x_0) = \frac{P(x_0 , x_1)}{P(x_0)} = k\binom{n_1}{x_1}[\{M\bar{x} + nx_0\}\{M\bar{x} + n(x_0 + 1)\} \cdots \cdot \{M\bar{x} + n(x_0 + x_1 - 1)\}][\{M(n - \bar{x}) + n(n_0 - x_0)\} \cdots \cdot \{M(n - \bar{x}) + n(n_0 + n_1 - x_0 - x_1 - 1)\}],$$

where

$$\frac{1}{k} = (M + n_0)(M + n_0 + 1) \cdots (M + n_0 + n_1 - 1)n^{n_1},$$

i.e., a hypergeometric distribution with parameters depending on x_0 . The bivariate distribution is thus not homoscedastic, but see below for a suggested method of dealing with this defect. The product moment correlation coefficient between x_0 and x_1 can be readily obtained from the variances of x_0 , x_1 , and x. Since the bivariate distribution is certainly not normal and in fact not even homoscedastic the value of this coefficient is not great; however, in this case it has some theoretical interest:

$$r_{(x_0, x_1)} = \sqrt{\frac{n_0 n_1}{(M + n_0)(M + n_1)}}.$$

If $n_0 = n_1$ then this value is equal to the Kuder-Richardson coefficient of the subtest of length n_0 . Thus the Kuder-Richardson coefficient has the further interpretation that it is the product moment correlation coefficient between the given test and a test of equal length and with similar items. However the bivariate distribution is of much more value in determining the range of scores likely to be obtained on a second test by people who all obtain the same score on the first test.

Because of their very skewed distributions, the two WM8 tests shown in Fig. 1 and described in Table 1 were chosen to illustrate the use that can be made of the results derived so far. These two tests are really the matched halves of a 60-item test, the items having been matched by Mollenkopf on difficulty and item-test correlation. The actual scatterplot between the two halves is shown in Fig. 2, the data having been arbitrarily grouped without reference to the observed frequencies (except to the observed mean and variance) in order to make possible the computation of a chi square. The numbers in parentheses are the theoretical frequencies obtained by fitting a bivariate negative hypergeometric (the theoretical frequencies were obtained from equation (2) by substituting the following values: $n = 60$, $\bar{x} = 13.51$, $r = .907$). The marginal distributions shown are the actual ones.

The chi square between theoretical and observed bivariate distributions has 16 degrees of freedom and is at about the 90-percent significance level. The fit is so good as to require consideration. The explanation presumably is that, because of the matching, the two marginal distributions are more similar than two randomly parallel halves would be, thus reducing the value of the chi square.

The Distribution of True Scores

So far the distribution of p in the population has remained arbitrary with the restriction that the regression of p on x should be linear and that

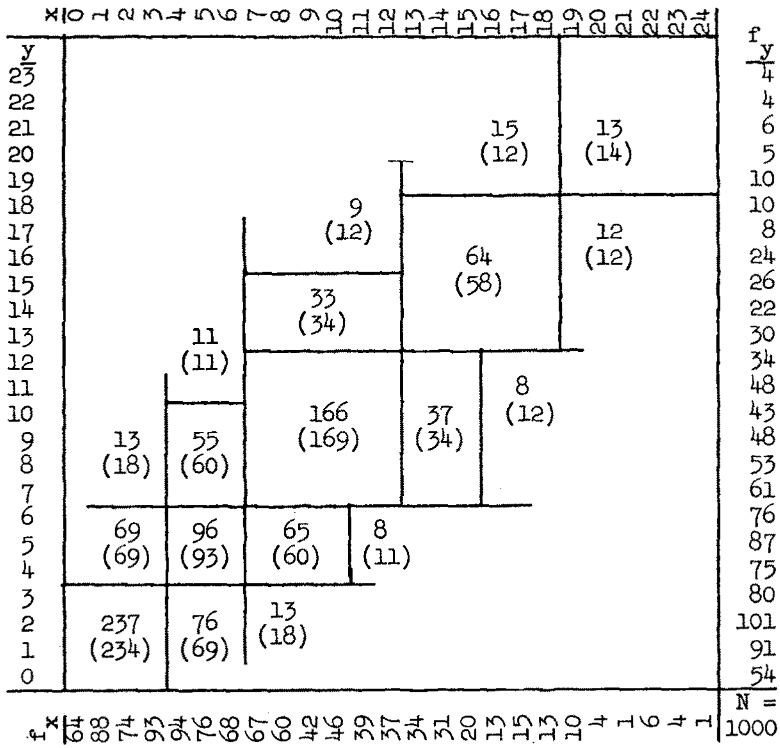

FIGURE 2

Bivariate Frequency Distribution of Two Parallel Tests, Showing Predicted Frequencies in Parentheses

p is continuously distributed in the range 0, 1. In general these restrictions do not determine the distribution of p uniquely, however it is known that if the distribution of p on x is linear, then the Pearson Type I or Beta distribution is a possible distribution for p. Furthermore, since the Beta distribution has at most one turning point and two points of inflection, it can be shown that all other solutions must have a large number (of the order n at least) of turning points or points of inflexion. Such distributions are unlikely to occur in practice, and as n becomes large it can be shown that their approximation by a Beta distribution becomes closer. It therefore seems reasonable to assume that the simplest distribution, i.e., the Beta distribution, is the correct one and proceed to develop the theory further on this assumption.

Let $f(p) = kp^l(1-p)^m$, where k is a constant and l and m are the parameters of this Beta distribution. Then

$$g(x) = \int_0^1 \binom{n}{x} p^x (1-p)^{n-x} k p^l (1-p)^m \, dp$$

$$= k\binom{n}{x}\int_0^1 p^{x+l}(1-p)^{m+n-x}\,dp$$

$= $ the right hand side of (1) with $M = l + m + 2$ and with $\dfrac{M\bar{x}}{n} = l + 1$.

Of great practical value is the possibility of writing down the likely distribution of values of p for persons who all obtain the same raw score. This is in effect the range of relative true scores for persons who all obtain the same raw score. What is required is the conditional distribution of p given x which can be found in the following way.

$$P(x, p) = \binom{n}{x}p^x q^{n-x} k p^l (1-p)^m$$

and

$$g(x) = \binom{n}{x}k\frac{\Gamma(l+x+1)\Gamma(m+n-x+1)}{\Gamma(l+m+n+2)}$$

but

$$P(p \mid x) = \frac{P(x, p)}{g(x)},$$

therefore,

$$P(p \mid x) = k' p^{l+x}(1-p)^{m+n-x},$$

where k' is a new constant.

Thus the conditional distribution of p is also a Beta function but with parameters depending on the raw score and the number of items as well as l and m. As the number of items increases, with corresponding increase in the possible values of x, the conditional distribution of p will have smaller and smaller variance and the distribution will become more and more concentrated about a point which approaches the value x/n. The value x/n as n becomes indefinitely large is usually referred to as the relative true score.

More precisely

$$\text{Mean}\ (p \mid x) = \frac{l + x + 1}{l + m + n + 2} \to x/n$$

as x and n become large,

$$\text{Var}\ (p \mid x) = \frac{\text{Mean}\ (p \mid x)}{l + m + n + 3}\,[1 - \text{Mean}\ (p \mid x)] \to 0$$

as x and n become large.

From the variance of the distribution of $p \mid x$ it is possible to deduce the product moment correlation between p and x. Thus

$$r^2_{(p,x)} = 1 - \sum_x \frac{\text{Var}\ (p \mid x)}{\text{Var}\ (p)}\, g(x).$$

The right-hand side of this equation simplifies to $n/(M + n)$ which is the Kuder-Richardson formula. Hence the product moment correlation coefficient between raw score and p is equal to the square root of the Kuder-Richardson coefficient. This result is not new, but its derivation is very simple here. The lack of homoscedasticity of the bivariate distribution makes the result of doubtful value, and although this difficulty may be removed at least to a considerable extent, it is probably just as simple and certainly more exact to work with the bivariate distribution.

To illustrate one of the applications of the derivations from the assumption of a Beta distribution of true scores, the 95-percent confidence limits of relative true scores were estimated for each raw score on a 20-item test administered to 100 children. These are tabulated in Table 2, values are found to the nearest .05 and were obtained from [7].

Corrections for Lack of Homoscedasticity

It has been noticed that the various bivariate surfaces studied are not homoscedastic, i.e., that the conditional distributions have variances which are not constant from one conditional distribution to the next. However in all the cases noted this lack of homoscedasticity can be overcome at least partially because the change in variance is solely due to a change of mean, i.e., Var $= k$ Mean $(1 -$ Mean) where k is constant for all possible conditional distributions in the bivariate surface. Under these conditions Kendall [2] has shown that it can be confidently expected that the transformation $u = \arcsin \sqrt{v}$ will give some improvement. In the case of $P(p \mid x)$ the transformation would be $u = \arcsin \sqrt{p}$, which would lead to a nearly constant column variance of $1/(1 + m + n + 3)$ and there would be similar

TABLE 2

Five-Percent Confidence Limits of Relative True Scores

Raw score	Range of relative true scores	Raw score	Range of relative true scores
0	0–.15	11	.40–.80
1	0–.15	12	.45–.80
2	0–.25	13	.50–.85
3	0–.30	14	.55–.90
4	.05–.35	15	.60–.90
5	.10–.40	16	.65–.95
6	.10–.45	17	.70–1.00
7	.15–.50	18	.75–1.00
8	.20–.55	19	.85–1.00
9	.20–.60	20	.85–1.00
10	.30–.70		

transformations for x, x_0, and x_1 with variances $(M + n)/n(M + 1)$, $(M + n_0 + n_1)/n_0(M + n_1 + 1)$, etc.

The correction for lack of homoscedasticity affects the values of the product moment correlation coefficients derived above. By similar methods to those used above it can be shown that, if x_0' and x_1' are the transformed values of x_0 and x_1,

$$r_{(x_0',x_1')} = \sqrt{\frac{n_0(n_1 - 1)}{(M + n_1)(M + n_0 + 1)}},$$

which is slightly different from the Kuder-Richardson formula. The correlation between probability or relative true score and raw score becomes

$$r_{(p',x')} = \sqrt{\frac{n}{n + M + 1}},$$

which is slightly different from the square root of the Kuder-Richardson formula. However the Kuder-Richardson formula itself remains invariant under the transformation if it is defined as the ratio of the variance of relative raw score to the variance of relative true score.

Summary

Assumptions

1. The distribution of the raw scores, x, can be represented by

$$g(x) = \int_0^1 \binom{n}{x} p^x(1 - p)^{n-x} f(p)\, dp.$$

2. The regression of the proportion-correct true score, p, on x is linear.

Deductions

1. The distribution $g(x)$ is the negative hypergeometric distribution.
2. The bivariate distribution of randomly parallel tests is the bivariate negative hypergeometric distribution.
3. The Kuder-Richardson formula 21 has a unique interpretation in terms of a parameter of the negative hypergeometric distribution, and the Spearman-Brown formula applies to it.

Further Assumption

1. The frequency distribution $f(p)$ is continuous.

Deduction

1. The only "reasonable" distribution function for p is the Pearson Type I or Beta function.
2. Confidence limits of p for each value of x can be calculated or obtained from Pearson's Tables of the Beta Function.

The empirical results presented here and in the references cited suggest that, if the usual checks for goodness of fit are applied, the theory presented will have many applications to practical problems of test construction and interpretation.

REFERENCES

[1] Keats, J. A. *A statistical theory of objective test scores.* Melbourne: Australian Council for Educational Research, 1951.

[2] Kendall, M. G. and Stuart, A. *The advanced theory of statistics.* New York: Hafner, 1958.

[3] Lord, F. M. Problems in mental test theory arising from errors of measurement. *J. Amer. statist. Ass.*, 1959, **54**, 472–479.

[4] Lord, F. M. An approach to mental test theory. *Psychometrika*, 1959, **24**, 283–302.

[5] Lord, F. M. A survey of observed test-score distributions with respect to skewness and kurtosis. *Educ. psychol. Measmt*, 1955, **15**, 383–389.

[6] Mollenkopf, W. G. Variation of the standard error of measurement. *Psychometrika*, 1949, **14**, 189–230.

[7] Pearson, K. (Ed.) Tables of the incomplete beta-function. *Biometrika Publications* London: University College, 1934.

[8] Raiffa, H. and Schlaifer, R. *Applied statistical decision theory.* Boston: Harvard Business School, 1961.

Manuscript received 4/14/61

Revised manuscript received 9/18/61