Nederlands Tijdschrift voor de Psychologie 29, 409-414

A BAYESIAN FORMULA SCORE FOR THE SIMPLE KNOWLEDGE OR RANDOM GUESSING MODEL

Ad H.G.S. van der VEN
Psychological Laboratory, University of Nijmegen, Nijmegen

Some disadvantages of the formula score, a well-known correction-for-guessing formula are discussed. A main disadvantage of this formula is that it is based on the regression of the number of items right on the number of items known.

A new formula score is proposed, which is derived from the regression of the number of items known on the number of items right. A uniform distribution is used as a prior distribution for the number of items known. In this manner each subject is treated in the same way, insofar as one's prior belief about the subject is the same for all subjects.

Introduction

Recently Van Naerssen (1972) again put forward the problem of estimating in achievement tests the true score (or true aptitude) given the observed score. In this paper an attempt is made to attack the problem by using a Bayesian approach.

In their simple knowledge or random guessing model for achievement tests Lord and Novick supposed "...that if an examinee knows the answer to an item he encounters, he gives that answer, and that if he does not know the answer, he... guesses at random" (Lord and Novick 1968: p. 303 and sections 14.1 through 14.4). They give the marginal distribution of X, the number of items right, as

$$P(X = x) = (1 - p)^{n-x} \sum_{\kappa=0}^{n} \binom{n-x}{x-\kappa} p^{x-\kappa} P(K = \kappa),$$

where

n = the number of items in the test, $n = 0, 1, 2, \ldots$

k = the number of items known, $K = 0, 1, 2, \ldots, n$

x = the number of items right, $x = k, k + 1, \ldots, n$

p = the probability of correctly guessing an item which is not known, $0 < p < 1$.

Capital letters are used to indicate stochastic variables; lower case letters indicate the specific values these stochastic variables may assume.

A well=known correction-for-guessing formula derived from this model, is the so-called *formula score* (see e.g. Lord and Novick 1968: formula 14.3.4, p. 306):

$$\hat{\kappa} = \frac{x - np}{1 - p}$$

Generally, p is set equal to $1/A$, where A is the number of alternatives. As one may easily see the estimate k may assume negative values. However, a more important disadvantage of this formula is, that it is based on $E(X \mid K = k)$ the regression of X on k, where it should be based on $E(K \mid X = x)$, the regression of K on x. The error which is made by starting from the regression of X on k is as follows. Generally, one may obtain

$$E(X \mid K = \kappa) = \kappa + p\,(n - \kappa)$$

Rewriting leads to

$$\kappa = \frac{E(X \mid K = \kappa) - np}{1 - p}$$

The error is made by substituting x for $E(X \mid K = k)$. An alternative procedure would be desired.

In this paper an attempt is made to establish an alternative formula score based on the regression of K on x. The problem is to estimate the number of items known, given a specific number of correct items. The simple formula score is not appropriate for it may give negative values. The method presented here uses the conditional distribution of K for a given observed score x: $P(K = x \mid X = x)$. Using Bayes' law one may obtain

$$P(K = \kappa \mid X = x) = \frac{P(X = x \mid K = \kappa)P(K = \kappa)}{\sum_{\kappa=0}^{x} P(X = x \mid K = \kappa)P(K = \kappa)}$$

The conditional expectation $E(K \mid X = x)$ is used as a final estimate for the number of items known. In order to find a solution for $E(K \mid X = x)$ the distribution $P(K = k)$ must be specified. Up to now this is an unsolved problem as Lord and Novick state: "Unfortunately the psychometrician does not have exact values for $P(X = x)$, $x = 0, 1, ..., n$, but only the approximations represented by the observed frequencies in the sample of' examinees at hand. Substituting these approximations into (14.3.2) is likely to lead to negative estimates for some of the

$P(K = k)$ - an absurd result. Such negative values can be avoided by linear programming techniques, but the estimated $P(K = k)$ is still likely to be intolerably irregular. For the present, no entirely satisfactory methods seem to be available to deal with this problem" (Lord and Novick 1968: p. 305, ch. 14.3). But even if a solution could be found, then there remains a following problem. For each group of examinees $P(K = k)$ could be different. In that case, for different groups of examinees, one could obtain different estimates for the number of items known, given the same number of items right. This situation is undesirable in educational practice.

Method

Another procedure is suggested, which automatically avoids the two problems mentioned above which are as follows:
(1) the specification of $P(K = k)$ and
(2) the possibility of different estimates of K given the *same* value of x. One can apriori choose a uniform or rectangular distribution for $P(K = k)$. From Bayesian statistics it is known that the uniform distribution "...seems to be a good representation of a diffuse state of prior knowledge" (Hayes and Winkler 1970: p. 485, ch. 8.18), where the concept of diffuseness is circumscribed as follows: "Suppose that a statistician wants to assess a prior distribution in a situation where he has very little or no prior information. More specifically, his prior information is such that it is 'overwhelmed' by the sample information. Then it is said that the statistician has a diffuse, or *informationless*, state of prior information" (Hayes and Winkler 1970: p. 482, ch. 8.17). One can also state that in choosing a uniform distribution one's prior credibility or degree of belief is the same for each value of k. Using a uniform distribution for $P(K = k)$ the conditional expectation can be expressed as follows:

$$E(K \mid X = x) = \sum_{\kappa=0}^{x} \frac{\kappa \binom{n-\kappa}{x-\kappa} p^{x-\kappa} (1-p)^{n-x} c}{\sum_{\lambda=0}^{x} \binom{n-\lambda}{x-\lambda} p^{x-\lambda} (1-p)^{n-x} c},$$

where λ, $\kappa = 0, 1, 2, \ldots, x$, and c is an arbitrary constant.

Simplification leads to

$$E(K \mid X = x) = \frac{\Sigma_{\kappa=0}^{x} \kappa(n-\kappa)(n-\kappa-1)\ldots(x-\kappa+1)p^{x-\kappa}}{\Sigma_{\kappa=0}^{x}(n-\kappa)(n-\kappa-1)\ldots(x-\kappa+1)p^{x-\kappa}}.$$

This formula cannot be simplified further. So, tables* must be prepared for $E(K \mid X = x)$ for different values of n, say 1, ..., 50, and of p, say 1/2, ..., 1/5. The reader can gain a more intuitive understanding of the obtained formula score by studying figure 1. This figure shows the relationship between the observed proportion of items right, x/n, and the estimated proportion of items known, k/n, where $k = E(K \mid X = x)$.

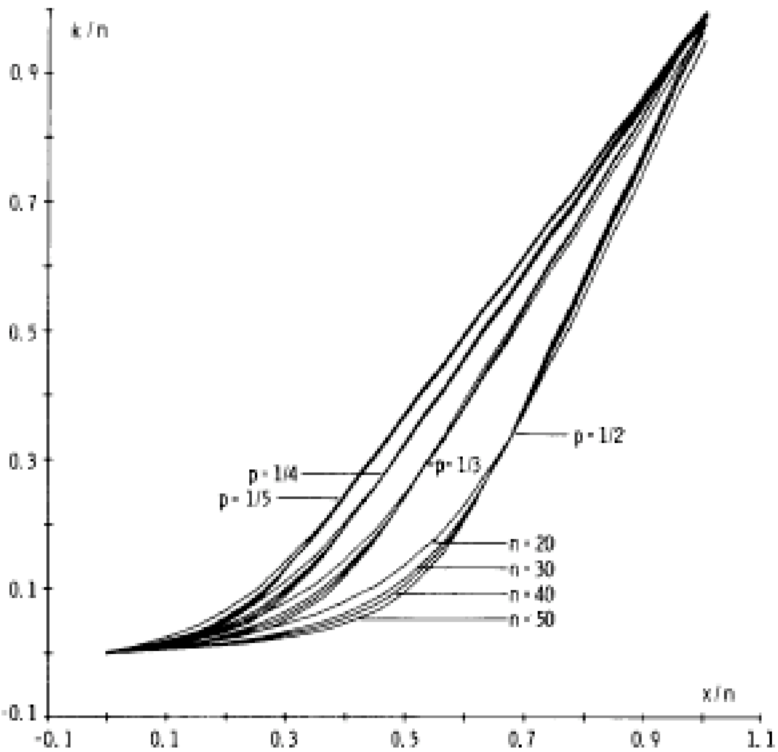

Fig. 1.

Curves are given for different values of p, p = 1/2, 1/3, 1/4 and 1/5, and for different values of n, n = 20, 30, 40, and 50. It is very obvious, that the differences between the curves mainly are accounted for by differences between p-values.

* An ALGOL program is available at the Department of Mathematical Psychology of the Psychological Laboratory of the University of Nijmegen, Erasmuslaan 16, Nijmegen.

Discussion

The method suggested here provides a solution for the problems mentioned above. The distribution of $P(K = k)$ is specified and reduces to a constant. For different groups the estimate of K is the same given the *same* value of x. Additionally, there is another advantage. Each subject is treated in the same way, insofar as one's prior belief about the subject is the same for all subjects.

The following conditions should hold for the Bayesian formula score to be applicable:

(1) The subject is in either one of two states, he either knows the answer or he does not.

(2) The probability of choosing the correct alternative is unity in case he knows the answer.

(3) In case he does not know the answer, all alternatives are equally likely.

The last condition holds, if all alternatives are equally attractive for subjects who do not know the answer. The use of distractors may be incompatible with condition (3) since for a specific subject some distractors may be more attractive than others. Condition (3) can easily be tested statistically by applying an appropriate test for goodness of fit to the response frequencies of the incorrect responses.

References

Hayes, W.L. and R.L. Winkler, 1971. Statistics: probability, inference and decision. New York: Holt, Rinehart and Winston.

Lord, F.M. and M.R. Novick, 1968. Statistical theories and mental test scores. Reading, Mass.: Addison-Wesley.

Naerssen, R.F. van, 1972. Het schalen van testscores. Ned. Tijdschr. Psychol. 27, 471 -485.

==