Eva Tesaříková

Probabilistic model of school-achievement test with double choice response: Variant I

Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica, Vol. 31 (1992), No. 1, 119--126

Persistent URL: http://dml.cz/dmlcz/120274

ACTA UNIVERSITATIS PALACKIANAE OLOMUCENSIS FACULTAS RERUM NATURALIUM

1992 Mathematica XXXI Vol. 105

PROBABILISTIC MODEL OF SCHOOL-ACHIEVEMENT TEST WITH DOUBLE CHOICE RESPONSE: VARIANT I

Eva Tesaříková

(Received March 10th 1991)

Abstract. Variant I of the probabilistic model of the school-achievement test with double-choice response presented in this paper is the first and simplest of the mathematical models describing the probabilistic structure of multiple-choice tests. The construction of this probabilistic model should take into account a certain difference between the real knowledge of the examined person and the result of the test evaluated by the examiner. One supposes among others that when the person under examination is really familiar with the tested topic then he selects both the correct responses from the offered alternatives.

Key words: school-achievement test with double-choice response, probabilistic model of school test.

MS Classification: 62P10, 62P15

Introduction

In 1978 the papers by S.Komenda and al. reported on the probabilistic model of school-achievement tests with a compulsory choice of one correct response among q offered alternatives. This simplest type of school-achievement test has been widely used in pedagogic practice but other tests have been employed as well. Recently multiple-choice tests have been introduced with compulsory choice of two or three correct responses among the offered alternatives, as well as association and sequence tests.

The aim of the present paper is to construct probabilistic models of the former types of school-achievement tests, i.e. tests with double-choice and triple-choice responses. In addition to the original models describing the probabilistic structure of the tests, attention has been paid to the determination of at least the statistical relations between the real knowledge of the tested person and the obtained results evaluated by the examiner. On the basis of this relations the rules for unbiased evaluation as well as for classification strategy can be stated.

The present paper is the first from a series of articles dealing with the described problems and reports on the simplest variant of the probabilistic model of school- achievement tests with double-choice response.

Assumptions of the model

Let us consider the school-achievement test with compulsory choice of two correct responses among $q>3$ response alternatives without any correction of conjectures so that the missing answer is evaluated as the incorect answer. The examined persons are informed about it before testing. It is also presumed that the test consists of n independent questions of the same difficulty, the number of offered alternatives is the same in all questions and that the function of the alternatives is equivalent in the case when double-choice response from all alternatives is conjuctured. Answers to one question should be chosen in order to avoid similarity and discrepancy.

The construction of the probabilistic model of such a test should take into account a certain difference between the real knowledge of the examined person and the result of the test evaluated by the examiner. The relative part of the topic to be examined which the tested person does not know can be expressed by the value of the parameter τ, while relative part of the real knowledge is expressed by $1-\tau$. When the difficulty of the individual questions is identical, the parameter τ also expresses the probability of the accurrence of the question within the person's unfamiliarity with the examined topic.

The probabilistic model of the double-choice test differs from the model of school-achievement test described elsewhere [1],[2] in the following three types of response of the examined person:

1) no correct answer
2) only one correct answer
3) both answers correct

According to the type of evaluation of the three types of responses, there are possible two variants of the probabilistic model of the school-achievement test with double choice response.

Model of probabilistic structure ot the test

Variant I of probabilistic model of the school-achievement test with double choice response corresponding to the above mentioned assumptions is derived from the presumption that when the person under examination is familiar with the tested topic he then selects both correct responses from all alternatives. If only one correct response is chosen then inaccurate knowledge of the tested branch or conjecture is suggested. This variant of the model appears to be important especially in those tests where a pair of correct responses represents a complex solution of the given question.

The description of the probabilistic structure of the test comprises the following notation for random events relative to i-th question of the test:

Z_i i-th question belongs to familiarity with the topic
N_i i-th question belongs to unfamiliarity with the topic

Random events Z_i , N_i form a complete system of events with probabilities

$$P(Z_i) = 1-\tau \quad , \quad P(N_i) = \tau$$

In relation to the registered results of the test the following three random events are considered

S_{i0} no correct answer was given to i-th question

S_{i1} only one correct answer was given to i-th question

S_{i2} both the answers were correct to i-th question

In the case of familiarity with the given topic the person examined will give the two correct responses. The following relations are in accordance with this presumption.

$$P(S_{i0}|Z_i) = 0 \ , \qquad P(S_{i1}|Z_i) = 0 \ , \qquad P(S_{i2}|Z_i) = 1 \ .$$

In the case of unfamiliarity with the topic tested by the question the examined person has to use the possibility of random choice. With regard to the presence of two correct responses among q>3 offered alternatives the following relations

$$P(S_{i0}|N_i) = \frac{\binom{q-2}{q}}{\binom{q}{2}} = \frac{(q-2)(q-3)}{q(q-1)}$$

$$P(S_{i1}|N_i) = \frac{\binom{q-2}{1}\binom{2}{1}}{\binom{q}{2}} = \frac{4(q-2)}{q(q-1)}$$

$$P(S_{i2}|N_i) = \frac{\binom{2}{2}}{\binom{q}{2}} = \frac{2}{q(q-1)}$$

hold for selection possibility in each question. These conditinal probabilities are defined by the hypergeometric distribution which is used just in the situation of dependent selection.

Unconditional probabilities of evets S_{i0} ,S_{i1} ,S_{i2} can be calculated according to the theorem of total probability:

$$p_{io} = P(S_{io}) = P(S_{io} \cap Z_i) + P(S_{io} \cap N_i) = \tag{1}$$

$$= P(Z_i)\, P(S_{io}|Z_i) + P(N_i)\, P(S_{io}|N_i) =$$

$$= \tau \frac{(q-2)(q-3)}{q(q-1)}$$

$$p_{i1} = P(S_{i1}) = P(S_{i1} \cap Z_i) + P(S_{i1} \cap N_i) = \tag{2}$$

$$= P(Z_i)\, P(S_{i1}|Z_i) + P(N_i)\, P(S_{i1}|N_i) =$$

$$= \tau \frac{4(q-2)}{q(q-1)}$$

$$p_{i2} = P(S_{i2}) = P(S_{i2} \cap Z_i) + PS_{i2} \cap N_i) = \tag{3}$$

$$= P(Z_i)\, P(S_{i2}|Z_i) + P(N_i)\, P(S_{i2}|N_i) =$$

$$= (1-\tau) + \tau \frac{2}{q(q-1)} = 1 - \tau\, (1 - \frac{2}{q(q-1)})$$

The assumption of the samė difficulty of questions and the same schema of offered answers allows us to omit the index of the question and to introduce the following notation

$$p_{io} = p_o\ ,\quad p_{i1} = p_1\ ,\quad p_{i2} = p_2\ ,\quad S_{io} = S_o\ ,\quad S_{i1} = S_1\ ,\quad S_{i2} = S_2$$

similarly as $Z_i = Z$, $N_i = N$ for all the considered i.

By an easy calculation

$$p_o + p_1 + p_2 = \tau \frac{(q-2)(q-3)}{q(q-1)} + \tau \frac{4(q-2)}{q(q-1)} + 1 - \tau \frac{q(q-1)-2}{q(q-1)} =$$

$$= \tau \frac{q^2 - 5q + 6 + 4q - 8 - q^2 + q + 2}{q(q-1)} + 1 = 1$$

is verified the fact that S_o, S_1, S_2 form a complete system of events.

Thus, each question of the test is related to the discrete probability space of elementary events with three elements S_o, S_1, S_2 the probability of which remains the same. The whole test, i.e. the whole series of n independent questions is related to discrete probability space which is a cartesian product of n particular probability spaces (S_0, S_1, S_2) in the sence of [3]. To this space can be associated three random variables M_0, M_1, M_2.

Variable M_0 represents the number of questions in the whole test to which no correct answer was given. This variable has a binomial probability distribution defined by the relation

$$P(M_0=m_0)=\binom{n}{m_0}\left(\tau\,\frac{(q-2)(q-3)}{q(q-1)}\right)^{m_0}\left(1-\tau\,\frac{(q-2)(q-3)}{q(q-1)}\right)^{n-m_0} \tag{4}$$

with the expectation

$$E(M_0) = n\,\tau\,\frac{(q-2)(q-3)}{q(q-1)} \tag{5}$$

and with the variance

$$D(M_0) = n\,\tau\,\frac{(q-2)(q-3)}{q(q-1)}\left(1-\tau\,\frac{(q-2)(q-3)}{q(q-1)}\right) \tag{6}$$

Variable M_1 represents the number of questions to which only one correct answer was given. This variable has a binomial probability distribution defined by relation

$$P(M_1=m_1) = \binom{n}{m_1}\left(\tau\,\frac{4(q-2)}{q(q-1)}\right)^{m_1}\left(1-\tau\,\frac{4(q-2)}{q(q-1)}\right)^{n-m_1} \tag{7}$$

with the expectation

$$E(M_1) = n\,\tau\,\frac{4(q-2)}{q(q-1)} \tag{8}$$

and with the variance

$$D(M_1) = n\,\tau\,\frac{4(q-2)}{q(q-1)}\left(1-\tau\,\frac{4(q-2)}{q(q-1)}\right) \tag{9}$$

The random variable M_2 representing the number of questions to which both correct responses were given has a binomial probability distribution in the form

$$P(M_2=m_2) = \binom{n}{m_2}\left(1 - \tau\,\frac{q(q-1)-2}{q(q-1)}\right)^{m_2}\left(\tau\,\frac{q(q-1)-2}{q(q-1)}\right)^{n-m_2} \tag{10}$$

with the expectation

$$E(M_2) = n\left(1 - \tau\,\frac{q(q-1)-2}{q(q-1)}\right) \tag{11}$$

and variance

$$D(M_2) = n\left(1 - \tau\,\frac{q(q-1)-2}{q(q-1)}\right)\left(\tau\,\frac{q(q-1)-2}{q(q-1)}\right) \tag{12}$$

In addition to the above three random variables it is useful to introduce the distribution of the random variable $X=M_0+M_1$ representing the number of questions to which both correct responses were not given at the same time. The distribution of this variable is binomial again

$$P(X=x) = \binom{n}{x}\left(\tau\,\frac{q(q-1)-2}{q(q-1)}\right)^{x}\left(1 - \tau\,\frac{q(q-1)-2}{q(q-1)}\right)^{n-x} \tag{13}$$

with expectation

$$E(X) = n\left(\tau\,\frac{q(q-1)-2}{q(q-1)}\right) \tag{14}$$

and variance

$$D(X) = n\left(\tau\,\frac{q(q-1)-2}{q(q-1)}\right)\left(1-\tau\,\frac{q(q-1)-2}{q(q-1)}\right) \tag{15}$$

With regard to the fact $m_0+ m_1+ m_2 = n$ the random vector $M=(M_0,M_1,M_2)$ has a multinomial probability distribution defined by the following relation

$$P(M_0=m_0, M_1=m_1, M_2=m_2) = \tag{16}$$

$$= \frac{n!}{m_0!m_1!m_2!}\left(\tau\,\frac{(q-2)(q-3)}{q(q-1)}\right)^{m_0}\left(\tau\,\frac{4(q-2)}{q(q-1)}\right)^{m_1}\left(1 - \tau\,\frac{q(q-1)-2}{q(q-1)}\right)^{m_2}$$

References

[1] Komenda, S.: Probabilistic model of the didactic test. Acta Univ. Palackianae Olomucensis, Fac.med. 85, 329-334 (1978 a)

[2] Komenda, S., Klementa, J.: Analýza náhodného v pedagogickém experimentu a praxi. SPN Praha 1981

[3] Tutubalin, V.N.: Teorie pravděpodobnosti. SNTL Praha 1978 Translation of russian original

Author's address: Department of Biophysics
Palacký University
Hněvotínská 3,
775 15 Olomouc
Czechoslovakia