Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica

Eva Tesaříková

Bayesian estimation of parameter τ in variant I of probabilistic model of double choice test

Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica, Vol. 34 (1995), No. 1, 175--178

Persistent URL: http://dml.cz/dmlcz/120327

Acta Univ. Palacki. Olomuc., Fac. rer. nat.,
Mathematica **34** (1995) 175–178

Bayesian Estimation of Parameter τ in Variant I of Probabilistic Model of Double Choice Test

Eva TESAŘÍKOVÁ

*Department of Algebra and Geometry, Faculty of Science,
Palacký University, Tomkova 40, 779 00 Olomouc, Czech Republic
E-mail: tesariko@risc.upol.cz*

(Received December 13, 1994)

Abstract

The paper follows the study of construction of probabilistic models of the school-achievement tests with double choice response and their statistical analysis published in 1992. The aim of this paper is a Bayesian estimation of the parameter τ representing the part of tested topic with which the examinee is unfamiliar on the basis of the realization of the test on a certain group of tested persons.

Key words: school-achievement test, double-choice response, probabilistic model, Bayesian estimation of the parameter of the test.

MS Classification: 62P10, 62P15

1 Introduction

This paper follows the study summarized in [1], [2], [3], and [4] where the probabilistic models of the school-achievement test with double choice response and their statistical analysis were given including certain simplifying assumptions.

Variant I of probabilistic model of such tests published by author of this paper in 1992 proceeds from the assumption that when the tested person is really familiar with the topic of the question, he or she will select both the correct responses among the offered alternatives. This variant suggests the expression of multinomic distribution of the random vector $M = (M_0, M_1, M_2)$ in the form

$$P(M_0 = m_0, M_1 = m_1, M_2 = m_2) = \tag{1}$$
$$= \frac{n!}{m_0! m_1! m_2!} \left(\tau \frac{(q-2)(q-3)}{q(q-1)}\right)^{m_0} \left(\tau \frac{4(q-2)}{q(q-1)}\right)^{m_1} \left(1 - \tau \frac{q(q-1)-2}{q(q-1)}\right)^{m_2}$$

where M_0, M_1, M_2 is a random variable expressing the number of questions of the test to which no correct response was given, only one correct response was given, both the correct responses were given by the examinee, respectively.

The distribution of the random vector $M = (M_0, M_1, M_2)$ of the registrated results of the test depends on one parameter τ which represents the proportion of the tested topic with which the examinee is unfamiliar.

2 Bayesian estimation of the parameter τ

Bayesian estimation including in this paper proceeds from the assumption of the realization of the double-choice test on the certain group of tested persons.

In case of investigated variant I of probabilistic model of such a test the probability distribution of registrated results of the test is given in the form (1). From this expression is derived the probability distribution of the sum $M_0 + M_1$, i.e. of the number of the questions to which the both correct answers were not given, as a binomial one. It depends on the parameter τ too, so we are able to see this probability distribution as a conditional one, i.e. as

$$P(M_0 + M_1 = m_0 + m_1 | \tau) = \tag{2}$$
$$= \binom{n}{m_0 + m_1} \left(\tau \frac{q(q-1)-2}{q(q-1)}\right)^{m_0+m_1} \left(1 - \tau \frac{q(q-1)-2}{q(q-1)}\right)^{n-m_0-m_1}$$

If we can consider the realization of the test on the certain group of examinees , it is reasonable to see the parameter τ as a random variable and consider the apriori probability density $f(\tau)$ of the parameter τ as a continuous function in the interval $\langle 0, 1 \rangle$.

The marginal probability distribution $P(M_0 + M_1 = m_0 + m_1)$ of the registered incorrect results in the test independent on the parameter τ then can be expressed by

$$P(M_0 + M_1 = m_0 + m_1) = \int_0^1 P(M_0 + M_1 = m_0 + m_1 | \tau)\, f(\tau)\, d\tau \tag{3}$$

It can be empirical realized as a distribution of the frequency of the numbers of the incorrect results in the test $N(m_0 + m_1)/N$ for $m_0 + m_1 = 0, 1, \ldots, n$ on the group of the N tested persons.

The aposteriori conditional probability density of the parameter τ in the subpopulation of the examinees with the number $m_0 + m_1$ of incorrect results in the test then can be expressed in the form

$$f(\tau | M_0 + M_1 = m_0 + m_1) = \frac{P(M_0 + M_1 = m_0 + m_1 | \tau)\, f(\tau)}{P(M_0 + M_1 = m_0 + m_1)} =$$

$$= \frac{f(\tau)\binom{n}{m_0+m_1}\left(\tau\frac{q(q-1)-2}{q(q-1)}\right)^{m_0+m_1}\left(1-\tau\frac{q(q-1)-2}{q(q-1)}\right)^{n-m_0-m_1}}{\displaystyle\int_0^1 P(M_0 + M_1 = m_0 + m_1 | \tau) f(\tau)\, d\tau} \tag{4}$$

Now denote by h the parametrical function

$$h(\tau) = \frac{\tau\dfrac{q(q-1)-2}{q(q-1)}}{1-\tau\dfrac{q(q-1)-2}{q(q-1)}} = \frac{\tau(q(q-1)-2)}{q(q-1)-\tau(q(q-1)-2)} \tag{5}$$

Formula (4) makes it possible to expresse the regression of this parametrical function in dependence on the number of incorrect results $m_0 + m_1$ in the test in the form

$$E\left(\frac{\tau(q(q-1)-2)}{q(q-1)-\tau(q(q-1)-2)} \,\middle|\, M_0 + M_1 = m_0 + m_1\right) =$$

$$= \int_0^1 \frac{\tau(q(q-1)-2)}{q(q-1)-\tau(q(q-1)-2)} f(\tau | M_0 + M_1 = m_0 + m_1)\, d\tau =$$

$$= \frac{1}{P(M_0 + M_1 = m_0 + m_1)} \frac{m_0 + m_1 + 1}{n - m_0 - m_1} \int_0^1 \binom{n}{m_0 + m_1 + 1} \cdot$$

$$\cdot \left(\tau\frac{q(q-1)-2}{q(q-1)}\right)^{m_0+m_1+1}\left(1-\tau\frac{q(q-1)-2}{q(q-1)}\right)^{n-m_0-m_1-1} f(\tau)\, d\tau$$

Using the formula (3) we get the relationship

$$E\left(\frac{\tau(q(q-1)-2)}{q(q-1)-\tau(q(q-1)-2)} \,\middle|\, M_0 + M_1 = m_0 + m_1\right) =$$

$$= \frac{m_0 + m_1 + 1}{n - m_0 - m_1} \frac{P(M_0 + M_1 = m_0 + m_1 + 1)}{P(M_0 + M_1 = m_0 + m_1)}.$$

According to the possibility of the estimation of the probability distribution $P(M_0 + M_1 = m_0 + m_1)$ by the empirical frequency distribution we obtain the empirical Bayesian estimator $\hat{h}(\tau)$ of the parametrical function $h(\tau)$ in the form

$$\hat{h}(\tau) = \frac{m_0 + m_1 + 1}{n - m_0 - m_1} \frac{N(m_0 + m_1 + 1)}{N(m_0 + m_1)} \qquad \text{for } m_0 + m_1 = 0, 1, \dots, n-1$$

So it is possible to obtain also an estimator $\hat{\tau}$ of the parameter τ in dependence on the value of $m_0 + m_1$ in the form

$$\hat{\tau}(m_0 + m_1) = \frac{\hat{h}(\tau)q(q-1)}{(1+\hat{h}(t))(q(q-1)-2)} =$$

$$= \frac{q(q-1)}{q(q-1)-2} \frac{\dfrac{m_0+m_1+1}{n-m_0-m_1}\dfrac{N(m_0+m_1+1)}{N(m_0+m_1)}}{1+\dfrac{m_0+m_1+1}{n-m_0-m_1}\dfrac{N(m_0+m_1+1)}{N(m_0+m_1)}} \tag{6}$$

Estimator (6) is usefull for every values of sum $m_0 + m_1$, expcept the case of $m_0 + m_1 = n$, on condition $\hat{\tau} \in \langle 0, 1 \rangle$. We can see that the Bayesian estimator $\hat{\tau}$ of the parameter τ takes into account the results of the tests given by whole the group of the other examinees.

References

[1] Tesaříková, E.: *Probabilistic model of school-achievement test with double-choice response—variant I.* Acta Univ. Palacki. Olomuc., Fac. rer. nat. **105**, Math. 31 (1992), 119–126.

[2] Tesaříková, E.: *Statistical analysis of probabilistic model—variant I of the school-achievement test with double choice response.* Acta Univ. Palacki. Olomuc., Fac. rer. nat. **105**, Math. 31 (1992), 126–147.

[3] Tesaříková, E.: *Variant II of probabilistic model of school-achievement test with double-choice response.* Acta Univ. Palacki. Olomuc., Fac. rer. nat. **110**, Math. 32 (1993), 159–164.

[4] Tesaříková, E.: *Statistical analysis of variant II of probabilistic model of the school-achievement test with double choice response.* Acta Univ. Palacki. Olomuc., Fac. rer. nat. **110**, Math. 32 (1993), 165–175.