FAIR MARKING OF MULTIPLE CHOICE TESTS

MANUEL L. ESQUÍVEL

This work is dedicated to Dominique Cellier and Etienne Ménard whom helped me to learn how to study Mathematics.

Abstract. In this note we state as a principle of fair marking that the expected value of the result of a multiple choice (MC) question answered completely at random must be zero. Under this principle there is only one way of marking a MC question in either one of two distinct ways of answering namely, indicating the false statement or indicating one or more of the true statements. We determine the probability of getting a pass classification on a ten item MC test for these two ways of answering showing a small advantage in the second form of answering.

1. Introduction

Multiple choice tests (MCT) may be considered as an interesting alternative for assessing knowledge and expertise particularly when computerized learning management systems are available allowing for automatic quiz proposals and correction. As opposed to more traditional methods for instance, open question exams, MCT are prone to misuse by guessing or more or less random answers. For that reason some techniques have been proposed to reduce the influence of guessing in the overall classification of a multiple choice test such as *formula scoring* (see [Frary] for a complete explanation), *liberal multiple choice tests* (see [Bush 99], [Bush 01]), or *permutational multiple choice questions* (as in [Farthing] and [Farthing 98]).

We present next some results showing that when a simple first principle of fair marking is adopted there is only one way of giving points to right and wrong choices in the most usual forms of MCT. Two types of MCT are studied, as an instance of a more general situation, in a choice of four alternatives when there are three correct statements and only one false statement. In the first one the allowed form of answering is to select the false statement. In the second one, one or more of the true statements are to be indicated. This second form of answering may be more effective to determine student knowledge if no information is given on the number of false statements.

2. Fair marking

Let us consider a multiple choice question with four possibilities to be answered. Among these four possibilities we know that three correspond to true statements and one corresponds to a false statement. Suppose that we want to mark answers. A natural question is how to weight each possibility in order to have a fair marking. There is one obvious first principle that we can consider.

Date: July 21, 2004.

1991 *Mathematics Subject Classification.* 97C40.

Our thanks to I. Cabral, A. Silva, T. Ferreira and J. T. Mexia with whom we discussed linear algebraic and statistical points related to this work. *This work was done with partial support from the FCT (Fundação para a Ciência e Tecnologia), program POCTI (Portugal/FEDER-EU)* .

Principle 2.1. The expected value of the mark of some examinee answering completely at random should be zero.

Using this principle we will present in the following some results according to the allowed forms of answering the question. The first result will show that this principle of fair marking implies the correction for guessing known as formula scoring.

Result 2.2. *Suppose you have a question with four possible answers with one, and only one of these four possible answers, being a false statement. If the* ***allowed form of answering is to fill one and only one box in the multiple choice question, namely the box corresponding to the false statement*** *then under the principle 2.1 of fair marking the correct answer has 1 mark and each of the three wrong ones has a -1/3 mark.*

Proof. In order to prove this first result just observe that there are $(s_i)_{i\in\{1,\ldots,4\}}$ possible answers described in the following table.

	s_1	s_2	s_3	s_4
T	x			
T		x		
T			x	
F				x

Let $\{m_1, m_2, m_3, m_4\}$ be the marks corresponding to each of the four possible answers. Principle 2.1 implies three conditions corresponding to the two following equations.

1. The marks corresponding to each one of the first three answers should be equal. This condition is expressed by
$$m_1 = m_2 = m_3$$
2. The correct answer s_4 should have full mark let us say $m_4 = 1$.
3. By an application of the principle 2.1 above, the sum of the marks corresponding to all the possible answers should be zero. To this condition corresponds the equation:
$$m_1 + m_2 + m_3 + m_4 = 0 \ ,$$
which is equivalent under the first condition to the equation:
$$3m_1 + m_4 = 0 \ ,$$
Now the second condition implies that $m_1 = m_2 = m_3 = -1/3$ as wanted.

□

Remark 2.3. This first result show that the principle of fair marking above implies the correction for guessing known as formula scoring. In fact, if formula scoring was applied to this situation we would have

$$FS = R - W/(C-1)$$

with FS the corrected formula score, R the number of items answered right, W the number of items answered wrong and C the number of choices per item. In this particular case we would have $FS = 0 - 1/(4-1) = -1/3$ just as stated in the result.

We now determine the probability of getting a pass classification on a test of ten independent questions answered, as in result 2.2 completely at random.

Remark 2.4. Consider the random variable X_1 describing the result of answering at random by signaling the false statement. According to the result we have that

$$X_1 = \begin{cases} 1 & \text{with probability } 0.25 \\ -1/3 & \text{with probability } 0.75 \ . \end{cases} \tag{2.1}$$

Such a random variable has mean value equal to zero and variance equal to 1/3. In order to evaluate the performance of a marking system one usually considers the likelihood of getting a pass mark on a test of ten independent questions marked according to the system under scrutiny (see [Farthing 98]. Let $(Y_j)_{j\in\{1,\dots,10\}}$ be a sequence of ten random variables equidistributed with X_1. The global mark received on a test of ten questions answered at random is given by $Z := \sum_{j=1}^{10} Y_j$. It is easy to see that the set of values taken by Z is given by the function

$$g(k) = k - \frac{10-k}{3}$$

where $k \in \{0,\dots,10\}$ represents the number of correct answers given to the ten questions of the test. Let us observe that the number of correct answers to a test, such as the one we are studying, is a binomial random variables with parameters $(10, 1/4)$. Supposing that a pass mark is obtained whenever the mark is equal or above 5/10 we will have that the probability of getting a pass mark is:

$$\sum_{k=7}^{10} \mathbb{P}[k=j] = \sum_{k=7}^{10} C_k^{10} 0.25^k 0.75^{10-k} = 0.00350571 \ ,$$

that is approximately 35/10000.

As a variation of the first result presented in result 2.2, we consider the somehow dual possibility of signaling (or not) the correct answers instead of signaling the false one. This allows for more latitude in the evaluation procedure as it is possible to differentiate attitudes towards knowing and not knowing and also to assess the confidence the examinee has in his/her own knowledge. The problem is not exactly dual of the one dealt with in the first result above as it is not necessary for the examinee to know that there is exactly one statement that is false.

Let us define first some notation. The following table describes the sequence $(s_i)_{i\in\{1,\dots,16\}}$ of all possible answers for a multiple choice question where there are three *true* statements and one *false* statement and the allowed form of answering is to fill all the boxes the examinee believes to correspond to the true answers.

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	s_{16}
T		x			x	x		x	x			x	x		x	
T			x		x		x	x		x		x		x	x	
T				x		x	x	x			x		x	x	x	
F									x	x	x	x	x	x	x	x

We suppose that for each one of the possible answers there are four possible marks denoted by $\{m_1, m_2, m_3, m_4\}$ corresponding to the following attitudes of the examinee in the answering procedure.

m_1 corresponds to putting a cross on a *true* statement;
m_2 corresponds to **not** putting a cross on a *true* statement;

m_3 corresponds to **not** putting a cross on a *false* statement;
m_4 corresponds to putting a cross on a *false* statement.

Being so the total mark corresponding to each set of answers s_i for each $i \in \{1, \ldots, 16\}$, is given by the linear combination of the marks m_1, m_2, m_3, m_4 that is detailed in the following set of formulas.

- s_1: $3m_2 + m_3$;
- s_2, s_3, s_4: $m_1 + 2m_2 + m_3$;
- s_5, s_6, s_7: $2m_1 + m_2 + m_3$;
- s_8: $3m_1 + m_3$;
- s_9, s_{10}, s_{11}: $m_1 + 2m_2 + m_4$;
- s_{12}, s_{13}, s_{14}: $2m_1 + m_2 + m_4$;
- s_{15}: $3m_1 + m_4$;
- s_{16}: $3m_2 + m_4$;

With these notations let us formulate the following result which shows that, under natural restrictive hypothesis there is only one solution for the set of total marks that will be given to each one of the answers (s_i) for $i \in \{1, \ldots, 16\}$ above.

Result 2.5. *Suppose that there are four statements, exactly one of which is false, and that the examinee is informed that there is at least one statement that is false. Suppose that the* ***allowed form of answering is to fill all the boxes the examinee considers to correspond to true statements****. Let as state as rules of marking the answers that:*

1. *Not answering at all will receive a mark equal to zero;*
2. *Filling all the boxes, which corresponds to the answer where all statements are considered to be true true, receives a mark equal to zero.*
3. *The only fully correct answer will receive full mark equal to one.*

Then in order to follow the principle 2.1 and the three rules stated above the only possible marks for each answer are:

- s_1*: 0;*
- s_2*,* s_3*,* s_4*: 1/3;*
- s_5*,* s_6*,* s_7*: 2/3;*
- s_8*: 1;*
- s_9*,* s_{10}*,* s_{11}*: -2/3;*
- s_{12}*,* s_{13}*,* s_{14}*: -1/3;*
- s_{15}*: 0;*
- s_{16}*: -1;*

Proof. The three conditions in the statement of the result have a correspondent translation into a set of equations satisfied by the marks of the answers we display next.

- Condition 1: $3m_2 + m_3 = 0$ and $3m_1 + m_4 = 0$
- Condition 2: $3m_1 + m_3 = 1$
- Condition 3: The sum of all marks should be zero or summing all the relations giving the marks for each possible answer: $3m_1 + 3m_2 + m_3 + m_4 = 0$

Observe that the matrix M correspondent to this system of equations

$$M = \begin{pmatrix} 0 & 3 & 1 & 0 \\ 3 & 0 & 0 & 1 \\ 3 & 0 & 1 & 0 \\ 3 & 3 & 1 & 1 \end{pmatrix}$$

has rank 3. Taking the variable m_1 as a parameter we then can solve the system of equations to get the solution $X = (m_1, m_2, m_3, m_4)$ of

$$M^{'}X = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} ,$$

the vector which components are:

$$m_1 \ , \ m_2 = \frac{-1}{3}(1 - 3m_1) \ , \ m_3 = 1 - 3m_1 \ , \ m_4 = -3m_1 \ .$$

Now, replacing these values in the formulas given above for the each of the marks of different answers s_i where $i = 1, \ldots, 16$ we obtain the result announced. □

As done previously we now determine the passing probability in a test of ten items answered completely at random.

Remark 2.6. Consider the random variable X_2 describing the result of answering at random by signaling the true statements in accordance with result 2.5. According to this result we have that:

$$X_2 = \begin{cases} -1 & \text{with probability } 1/16 \\ -2/3 & \text{with probability } 3/16 \\ -1/3 & \text{with probability } 3/16 \\ 0 & \text{with probability } 1/8 \\ 1/3 & \text{with probability } 3/16 \\ 2/3 & \text{with probability } 3/16 \\ 1 & \text{with probability } 1/16 \ . \end{cases} \tag{2.2}$$

Such a random variable has mean value equal to zero and variance equal to 1/3, that is the same location and dispersion parameters as the random variable X_1 defined in the remark 2.4 correspondent to signaling only the false answer in accordance with result 2.2. Let us determine the probability of getting a pass mark (5/10 or above) in a examination with ten multiple choice questions as in the result 2.5. Consider $(Y_j)_{j\in\{1,\ldots,10\}}$, a sequence of ten random variables equidistributed with X_2. The global mark received on a test of ten questions answered at random is given by $Z := \sum_{j=1}^{10} Y_j$. In order to determine the law of Z we will use pseudo-probability generation function technique. We define the pseudo-probability generating function of X_2 as:

$$\begin{aligned} \phi_{X_2}(a_1, a_2, a_3, a_4, a_5, a_6, a_7) =& \frac{1}{16}t^{a_1} + \frac{3}{16}t^{a_2} + + \frac{3}{16}t^{a_3} + \frac{2}{16}t^{a_4} + \\ &+ \frac{3}{16}t^{a_5} + \frac{3}{16}t^{a_6} + \frac{1}{16}t^{a_7} \ , \end{aligned} \tag{2.3}$$

where we know that for $t > 0$

$$\phi_{X_2}(-1, -\frac{2}{3}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3}, 1)$$

is, only formally, the probability generating function of X_2. By the mutual independence of the random variables $(Y_j)_{j\in\{1,\ldots,10\}}$ the pseudo-probability generating function of Z is given by:

$$\phi_Z(a_1, a_2, a_3, a_4, a_5, a_6, a_7) = \phi_{X_2}(a_1, a_2, a_3, a_4, a_5, a_6, a_7)^{10} \ .$$

Using Mathematica *TM*, the full expression of ϕ_Z taken at the vector $(-1, -\frac{2}{3}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3}, 1)$ is obtained as

$$\phi_Z(-1, -\frac{2}{3}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3}, 1) = \sum_{k=0}^{60} p_k t^{\frac{k}{3}-10} \ ,$$

where $(p_k)_{k\in\{1,\ldots,60\}}$ represent the probabilities associated to the different values $(k/3 - 10)_{k\in\{1,\ldots,60\}}$ that the random variable Z can take. Summing the probabilities correspondent to the values of Z that are greater than $5 = 45/3 - 10$, we get that:

$$\mathbb{P}[Z \geq 5] = \sum_{k=45}^{60} p_k = \frac{2714368909}{1099511627776} = 0.0024687 \ ,$$

that is, approximately, 25/10000. This result is naturally better than the result obtained for marking only the false statements.

3. Conclusion

MCQ tests with a stem, a false statement and three correct statements may be answered at least in two different ways. In one of these ways the allowed form of answering is to indicate what the examinee thinks to be the true statements and not the false one. Under the fair marking principle for which the expected value of the mark of some examinee answering completely at random is zero, there is only one way of marking each one of the possible answers. Again, under this fair marking principle the probability of getting a pass mark in a MCQ test with ten questions of the type described is $25/10,000$ which is better than the correspondent probability when the form of answering is to indicate the false statement.

References

[Bush 99] M. Bush, Alternative marking schemes for on-line multiple choice tests, *7th Annual Conference on the teaching of computing*, Belfast, 1999.

[Bush 01] M. Bush, A Multiple choice test that rewards partial knowledge, *Journal of Further and Higher Education*, Vol. 25, No. 2 2001.

[Farthing] D. W. Farthing, D. McPhee, Multiple choice for honours-level students? A statistical evaluation, preprint.

[Farthing 98] D. W. Farthing, D. M. Jones, D. McPhee, Permutational multiple-choice questions: an objective and efficient alternative to essay-type examination questions, *ITiCSE'98* Dublin, Ireland.

[Frary] R. B. Frary, Formula Scoring of Multiple-Choice Tests (Corrections for Guessing), *Instructional Topics in educational Measurement*.

Departamento de Matemática, FCT/UNL, Quinta da Torre, 2829-516 Caparica, Portugal
E-mail address: mlefct.unl.pt
URL: http://ferrari.dmat.fct.unl.pt/personal/mle