Multiple Choice Grading

a topic in

DM534 – Introduction to Computer Science

Kim Skak Larsen

Department of Mathematics and Computer Science (IMADA)
University of Southern Denmark (SDU)

kslarsen@imada.sdu.dk

November 24, 2016

Multiple Choice Grading

We want to automate the boring process of grading multiple choice exams!

This is how computer scientists think:

When you do something for the second time:
"it must be possible to automate this!"

This is easy if you know how to assign a score to right as well as wrong answers to each question.

Now, we recursively arrive at the box above again!

Multiple Choice Assumptions

We assume that a question has

- k options, and
- exactly one correct answer.

Multiple Choice Assumptions

We assume that a question has

- k options, and
- exactly one correct answer.

Not like this:

What is $1 + 1$?

- ☐ This is not the answer.
- ☐ Not this one either...
- ☐ 2.
- ☐ You went too far – back up!

Multiple Choice Assumptions

We assume that a question has

- k options, and
- exactly one correct answer.

More like this:

What is $\sum_{i=0}^{\infty} \frac{1}{2^i}$?

- ☐ 1
- ☐ ∞
- ☐ 2

Spoiler!

There is a *unique, fair* way of grading (up to a linear transformation).

This means that once we have decided, for instance, that

- a perfect set is 100, and
- a blank set is 0,

then there is a unique, fair way of grading!

We now focus on grading one question in a fair manner.

Axiom "Emptiness"

Which score should we give to a blank question (no checked boxes)?

What is $\sum_{i=0}^{\infty} \frac{1}{2^i}$?

- ☐ 1
- ☐ ∞
- ☐ 2

Axiom "Emptiness"

Which score should we give to a blank question (no checked boxes)?

What is $\Sigma_{i=0}^{\infty} \frac{1}{2^i}$?

- ☐ 1
- ☐ ∞
- ☐ 2

Axiom ("Emptiness")

If no boxes are checked, the score is zero.

Partial Credit

In classic written (and oral) exams, we give a lot of *partial credit*, i.e., credit for knowing *something* though it's less than knowing *everything* about a concrete issue.

Imagine we want to do the same here and give some (not full) credit for "catching" the correct answer with strictly fewer than all boxes checked. For instance as in

What is $\sum_{i=0}^{\infty} \frac{1}{2^i}$?

- ☒ 1
- ☐ ∞
- ☒ 2

Axiom "Fullness"

Which score should we give to a fully checked question (all boxes checked)?

What is $\sum_{i=0}^{\infty} \frac{1}{2^i}$?

☒ 1

☒ ∞

☒ 2

Axiom "Fullness"

Which score should we give to a fully checked question (all boxes checked)?

What is $\sum_{i=0}^{\infty} \frac{1}{2^i}$?

☒ 1

☒ ∞

☒ 2

Axiom ("Fullness")

If all boxes are checked, the score is zero.

Terminology

- $k \geq 2$ options
- a checked boxes, $0 \leq a \leq k$
- $R(k, a)$ – the score when the correct answer is checked ("R" for "right")
- $W(k, a)$ – the score when the correct answer is *not* checked ("W" for "wrong")
- When discussing more than one question at a time, we use indexes $k_1, k_2, \ldots$ and $a_1, a_2, \ldots$

Axiom "Positivity"

What can we say about the value $R(k, 1)$?

Axiom "Positivity"

What can we say about the value $R(k,1)$?

Axiom ("Positivity")

$R(k,1) > 0$.

On Measuring Knowledge Transfer

- We can use the partial credit idea to define another axiom, which we can use to derive a scoring function through a thought experiment.
- We remark that this is independent of whether or not we later allow students to check more than one box per question or not.
- What should the *combined* credit for the first to questions on the next slide be compared with the third?

What is $\sum_{i=0}^{\infty} \frac{1}{2^i}$?

- ☒ 1
- ☐ ∞
- ☒ 2

Is 42 prime?

- ☐ Yes
- ☒ No

What is $\sum_{i=0}^{\infty} \frac{1}{2^i}$ and is 42 prime?

- ☐ 1 and yes
- ☐ ∞ and yes
- ☐ 2 and yes
- ☒ 1 and no
- ☐ ∞ and no
- ☒ 2 and no

Axiom "Composition"

As observed on the previous slide, the knowledge required is identical in the two situations, so the score should be the same.

Axiom ("Composition")

$$R(k_1 k_2, a_1 a_2) = R(k_1, a_1) + R(k_2, a_2)$$

Partial Credit Scoring

From the Axiom "Composition",

$$R(k_1 k_2, a_1 a_2) = R(k_1, a_1) + R(k_2, a_2)$$

Letting $k_2 = 1$ and $a_1 = 1$, we get

$$R(k_1, a_2) = R(k_1, 1) + R(1, a_2)$$

Since R is symmetric in the two arguments, if we can establish what kind of function $R(k, 1)$ is, $R(1, a)$ must also be such a function.

Only considering non-negative integer arguments,

$$R(k_1 k_2, 1) = R(k_1, 1) + R(k_2, 1)$$

This can be seen as a function of only the first argument, so we just need to determine what kind of function fulfills

$$f(xy) = f(x) + f(y)$$

Additivity Claim

Lemma

If g is continuous, $g(x+y) = g(x) + g(y)$, and $g(1) = c$, then $g(x) = cx$.

Proof.

- By induction, we can show that it holds for all integers.
- By induction, we can show that $g(\frac{1}{x}) = c\frac{1}{x}$ for all integers x.
- By induction, we can show that $g(\frac{x}{y}) = c\frac{x}{y}$ for all integers x, y.
- Since the rational numbers are dense and g is continuous, a proof by contradiction can establish that it holds for all real numbers.

□

The Form of f

Lemma

If $f : \mathbb{R}_+ \to \mathbb{R}$ and $f(xy) = f(x) + f(y)$, then $f(x) = c \ln x$ for some constant c.

Proof.

Define $g(x) = f(e^x)$. Then

$$g(x+y) = f(e^{x+y}) = f(e^x e^y) = f(e^x) + f(e^y) = g(x) + g(y)$$

By the additivity claim, $g(x) = cx$ (for $c = g(1) = f(e)$), so $f(e^x) = cx$.

If $y = e^x$, then $x = \ln y$, so substituting into $f(e^x) = cx$, we get

$$f(y) = c \ln y$$

□

Back to the Scoring Function

So,

$$R(k,a) = c \ln k + c' \ln a, \text{ for some } c, c'.$$

By Axiom "Fullness", $R(k,k) = 0$, so

$$\begin{aligned} & c \ln k + c' \ln k = 0 \\ \Downarrow & \\ & c \ln k = -c' \ln k \\ \Downarrow & \\ & c = -c' \end{aligned}$$

Since $R(k,1) = c \ln k$, by Axiom "Positivity", we must have $c > 0$. Thus,

$$R(k,a) = c \ln k - c \ln a = c \ln \frac{k}{a}$$

Instead of multiplying all scores with a fixed arbitrary c, we can normalize using $c = 1$ and/or switch base of the logarithm (scaling). So,

$$R(k,a) = \log \frac{k}{a}$$

Guessing

An example multiple choice question:

My investment pay-off is 40 with probability $\frac{3}{4}$ and 48 with probability $\frac{1}{4}$.

What is my *expected* pay-off?

- ☐ 40.75
- ☐ 42
- ☐ 44
- ☐ 50

Axiom "Guessing"

- We cannot distinguish a lucky guess from a correct answer.
- We cannot distinguish an unlucky guess from a wrong answer.
- What do we want if a student guesses on lots of questions?

Axiom "Guessing"

- We cannot distinguish a lucky guess from a correct answer.
- We cannot distinguish an unlucky guess from a wrong answer.
- What do we want if a student guesses on lots of questions?

There is no knowledge transfer in guessing, so we want:

Axiom ("Guessing")

The expected value of guessing should be zero.

The Expected Value of Guessing

If one checks a boxes out of k at random, the expected gain should be zero.

The Expected Value of Guessing

If one checks a boxes out of k at random, the expected gain should be zero.

Event	#ways

The Expected Value of Guessing

If one checks a boxes out of k at random, the expected gain should be zero.

Event	#ways
place a check marks	

The Expected Value of Guessing

If one checks a boxes out of k at random, the expected gain should be zero.

Event	#ways
place a check marks	$\binom{k}{a}$

The Expected Value of Guessing

If one checks a boxes out of k at random, the expected gain should be zero.

Event	#ways
place a check marks	$\binom{k}{a}$
place a check marks, *not* including the correct answer	

The Expected Value of Guessing

If one checks a boxes out of k at random, the expected gain should be zero.

Event	#ways
place a check marks	$\binom{k}{a}$
place a check marks, *not* including the correct answer	$\binom{k-1}{a}$

The Expected Value of Guessing

If one checks a boxes out of k at random, the expected gain should be zero.

Event	#ways
place a check marks	$\binom{k}{a}$
place a check marks, *not* including the correct answer	$\binom{k-1}{a}$
place a check marks, *including* the correct answer	

The Expected Value of Guessing

If one checks a boxes out of k at random, the expected gain should be zero.

Event	#ways
place a check marks	$\binom{k}{a}$
place a check marks, *not* including the correct answer	$\binom{k-1}{a}$
place a check marks, *including* the correct answer	$\binom{k-1}{a-1}$

The Expected Value of Guessing

If one checks a boxes out of k at random, the expected gain should be zero.

Event	#ways
place a check marks	$\binom{k}{a}$
place a check marks, *not* including the correct answer	$\binom{k-1}{a}$
place a check marks, *including* the correct answer	$\binom{k-1}{a-1}$

We want

$$\frac{\binom{k-1}{a-1}}{\binom{k}{a}} R(k,a) + \frac{\binom{k-1}{a}}{\binom{k}{a}} W(k,a) = 0$$

The Expected Value of Guessing

If one checks a boxes out of k at random, the expected gain should be zero.

Event	#ways
place a check marks	$\binom{k}{a}$
place a check marks, *not* including the correct answer	$\binom{k-1}{a}$
place a check marks, *including* the correct answer	$\binom{k-1}{a-1}$

We want

$$\frac{\binom{k-1}{a-1}}{\binom{k}{a}} R(k,a) + \frac{\binom{k-1}{a}}{\binom{k}{a}} W(k,a) = 0$$

$$\Updownarrow$$

$$\binom{k-1}{a-1} R(k,a) + \binom{k-1}{a} W(k,a) = 0$$

The Expected Value of Scoring

$$\binom{k-1}{a-1} R(k,a) + \binom{k-1}{a} W(k,a) = 0$$

$$\Updownarrow$$

$$W(k,a) = \frac{-\binom{k-1}{a-1} R(k,a)}{\binom{k-1}{a}}$$

and

$$\frac{\binom{k-1}{a-1}}{\binom{k-1}{a}} = \frac{\frac{(k-1)!}{(a-1)!(k-a)!}}{\frac{(k-1)!}{a!(k-1-a)!}} = \frac{a!(k-1-a)!}{(a-1)!(k-a)!} = \frac{a}{k-a}$$

so

$$W(k,a) = -\frac{a}{k-a} \log \frac{k}{a}$$

All the Axioms and the Scoring Function

Axioms – Formal Formulation	
$W(k,0) = 0$	(emptiness)
$R(k,k) = 0$	(fullness)
$R(k,1) > 0$	(positivity)
$R(k_1 k_2, a_1 a_2) = R(k_1, a_1) + R(k_2, a_2)$	(composition)
$\binom{k-1}{a-1} R(k,a) + \binom{k-1}{a} W(k,a) = 0,\ a > 0$	(guessing)

A question with k options where a are checked should give the score

$$\begin{cases} 0, & \text{if } a = 0 \\ \log \frac{k}{a}, & \text{if } a > 0 \text{ and right answer} \\ -\frac{a}{k-a} \log \frac{k}{a}, & \text{if } a > 0 \text{ and wrong answer} \end{cases}$$

Scaling for an Exam Set

Choose your favorite logarithm function, log, for computing $\log \frac{k}{a}$.

For an exam set with n questions, the maximum score is then

$$m = \Sigma_{i=1}^{n} \log k_i$$

To get a max score of 100%, let $c = \frac{100}{m}$, and use

$$R(k, a) = c \log \frac{k}{a} \quad \text{and} \quad W(k, a) = -c \frac{a}{k-a} \log \frac{k}{a}$$

Implications

We showed that there was a *unique* choice, given the axioms.

Thus, if any different grading function is used, *at least one* fairness axiom is violated.

Most invent an ad-hoc (unfair) method.

Thus, ease of grading or ease of explaining is often given precedence over fairness.

Examples for Small k (scaled so $\log k = 1$)

a	0	1	2
Right		1	0
Wrong	0	-1	

a	0	1	2	3
Right		1	0.631	0
Wrong	0	-0.5	-1.262	

a	0	1	2	3	4
Right		1	0.5	0.208	0
Wrong	0	-0.333	-0.5	-0.623	

These tables *cannot* be used in the same test; scaling has to be the same globally and not separate for different k.

References

This talk was based on the following publication:

Gudmund Skovbjerg Frandsen and Michael I. Schwartzbach.
A singular choice for multiple choice.
SIGCSE Bulletin, 38(4):34–38, 2006.