5-1-2021

Shall We Ignore All Intermediate Grades?

Christian Servin
El Paso Community College, cservin1@epcc.edu

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:

Technical Report: UTEP-CS-21-54

Shall We Ignore All Intermediate Grades?

Christian Servin, Olga Kosheleva, and Vladik Kreinovich

Abstract In most European universities, the overall student's grade for a course is determined exclusively by this student's performance on the final exam. All intermediate grades – on homework, quizzes, and previous texts – are, in effect, ignored. This arrangement helps gauge the student's performance by the knowledge that the student shows at the end of the course. The main drawback of this approach is that some students do not start studying until later, thinking that they can catch up and even get an excellent grade – and this hurts their performance. To motivate students to study hard throughout the semester, most US universities estimate the overall grade for the course as a weighted average of the grade on the final exam and of all intermediate grades. In this paper, we show that even when a student is already motivated, to accurately gauge the student's level of knowledge it is important to take intermediate grades into account.

1 Formulation of the Problem

Two systems of grading. In most countries, in most universities, in most courses, students have intermediate tests and quizzes, homeworks, labs, most of which are graded. At the end of the course, there is usually a final exam which is also graded.

Christian Servin
Computer Science and Information Technology Systems Department
El Paso Community College (EPCC), 919 Hunter Dr., El Paso, TX 79915-1908, USA
e-mail: cservin1@epcc.edu

Olga Kosheleva
Department of Teacher Education, University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA, e-mail: olgak@utep.edu

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA, e-mail: vladik@utep.edu

In the US, usually, the overall grade for the course is estimated by combining the grade for the final exam and all intermediate grades – most frequently, by taking the weighted average. In contrast, in most European countries, the overall grade for the course is the grade on the final exam, with intermediate grades serving only on a pass-fail basis – to be able to take the final exam, a student needs to have a satisfactory average on all intermediate exams.

Why the difference: pro and contra. The ultimate goal of the grade is that it should reflect the knowledge that the student acquired after taking the course. From this viewpoint, it seems to make sense to use the European system: if the student did not do perfectly well on the intermediate exams, but eventually learned the material perfectly, this student should get a perfect grade.

The downside of this approach is that some students procrastinate and only start studying much later, thinking that they still have a chance to learn the material and get the perfect grade. Sometimes they succeed, but often they don't: they get a low grade or even fail the class. This problem did not bother people in the past, when a relatively small number of people could get higher education. Students were accepted only after very competitive final exams, so if a student does not want to study hard, well, good riddance, there are plenty of practically as good students eager to take this student's place.

However, nowadays, when jobs not requiring education are more and more automated, societies need highly educated people to survive in the global competition. Universities accept a large number of people, and not all of them are prepared to work hard. So we need to motivate them to study – and the US system clearly motivates students to start studying from the very beginning, since otherwise their not so good intermediate grades will affect their final grade for the class.

From this motivational viewpoint, some version of a US system is preferred.

What we do in this paper. In this paper, we show – somewhat unexpectedly – that even if we have perfect motivations, and we are willing to gauge a student by the knowledge he/she attained after the course, we still need to take intermediate grades into account.

2 Analysis of the Problem and the Resulting Conclusion

Main idea. A typical US final exam lasts for 2 hours and 45 minutes. The final exam is supposed to be comprehensive, covering all main topics that were studied in the course. It is possible to cover many things in this time, but clearly not everything that was taught during the semester.

If a student correctly solved 7 problems out of 10, but did not do well on intermediate assignments, then maybe this student's degree of knowledge is less than 70%, he/she just got lucky by the fact that questions on the exam – which were reasonably randomly selected from all possible questions – were mostly from the parts of the material that this student knew. On the other hand, if, in addition to correctly solving

7 problems out of 10 on the final exam, the student also had a similar satisfactory grade for all intermediate assignments, we are much more confident that this student indeed knows at least 70% of the material.

Thus, to accurately gauge the student's degree of knowledge, it is necessary to also take into account this student's intermediate grades.

How to take intermediate grades into account? The above qualitative argument shows that it is desirable to take intermediate grades into account. A natural next question is how exactly to take the intermediate grades into account when computing the overall grade for the course.

How the grade on the final exam is usually computed. To answer this question, let us first recall how the grade on the final exam is usually computed.

Each question on the final exam usually consists of several parts (explicit or implicit "sub-questions"), and the grade for this question is determined by how many of these parts the student answered correctly. For example, if an assignment is to apply a multi-stage algorithm, the instructor will check whether each of the steps is correctly performed.

The grade for the final exam is then obtained by adding the grades for all the questions. From this viewpoint, the grade for the final exam is determined by the number of correctly answered sub-questions.

In general, if on the final grade, out of s sub-questions, the student correctly answered n of them, then the student gets a fraction

$$\widetilde{p} = \frac{n}{s} \tag{1}$$

of the maximum possible score.

What do we want to estimate and how can we estimate it based on the final exam. A natural measure of the student's knowledge is the proportion p of all possible sub-questions to which the student knows the correct answer. This means that for each randomly selected sub-question, the probability that the student knows the correct answer to this sub-question is equal to the proportion p.

The actual number n of sub-questions that the student answered correctly on the final exam can be obtained by adding s independent 0-1-valued random variables v_i describing whether the i-th sub-question was answered correctly. For each of these variables, v_i the mean value is equal to

$$E[v_i] = 1 \cdot p + 0 \cdot (1-p) = p, \tag{2}$$

and the variance is equal to

$$E[(v_i - E[v_i])^2] = p \cdot (1-p)^2 + (1-p) \cdot (0-p)^2 = p \cdot (1-p) \cdot (1-p+p) =$$
$$p \cdot (1-p); \tag{3}$$

see, e.g., [1] for this and following formulas.

For the sum n of several independent random variables, the mean is equal to the sum of the means, and the variance is equal to the sum of the variances, so $E[n] = p \cdot s$ and $V[n] = p \cdot (1-p) \cdot s$.

During the 2 hours and 45 minutes we can ask a lot of sub-questions, so the number s is reasonably large. It is known that the probability distribution of the sum of a large number of small independent random variables is close to Gaussian – this is a consequence of the Central Limit Theorem (and the main reason why normal distributions are ubiquitous). Thus, we can conclude that the number n of correctly answered sub-questions is normally distributed, with mean $E[n] = p \cdot s$ and standard deviation $\sigma[n] = \sqrt{p \cdot (1-p) \cdot s}$.

The difference $n - E[n] = n - p \cdot s$ is normally distributed, with 0 mean and standard deviation $\sigma = \sqrt{p \cdot (1-p) \cdot s}$. For large n, the difference $n - p \cdot s$ is small, so $n \approx p \cdot s$ and thus,

$$p \approx \widetilde{p} \stackrel{\text{def}}{=} \frac{n}{s}, \tag{4}$$

hence $\sigma[n] \approx \widetilde{\sigma} \stackrel{\text{def}}{=} \sqrt{\widetilde{p} \cdot (1-\widetilde{p}) \cdot s}$. So, once we know the number n of sub-questions that the student has correctly answered on the final exam, we can conclude that the value $p \cdot s$ is normally distributed with mean n and standard deviation $\widetilde{\sigma}$.

Thus, the actual (unknown) grade p is also normally distributed, with mean

$$\widetilde{p} = \frac{n}{s} \tag{5}$$

and standard deviation

$$\sigma \approx \sqrt{\frac{\widetilde{p} \cdot (1-\widetilde{p})}{s}}. \tag{6}$$

How to take into account intermediate grades: idea. In addition to the s sub-questions that form the final exam, the student also answered several sub-questions before that, as part of intermediate tests, quizzed, homeworks, etc. Let us denote the overall number of such sub-questions by S, and the overall number of those of these sub-questions that the student answered correctly by N.

This does not necessarily mean that this is how much the student knows now, at the time of the final exam: the student may have learned what he or she missed earlier. What we can conclude, however, is that the student's degree of knowledge is at least as large as what can be concluded from this student's intermediate grades.

In the worst case scenario, when the student did not learn anything since previous exams, this student's degree of knowledge is normally distributed with the mean

$$\widetilde{P} = \frac{N}{S} \tag{7}$$

and standard deviation

$$\Sigma = \sqrt{\frac{\widetilde{P} \cdot (1-\widetilde{P})}{S}}. \tag{8}$$

Thus, with high certainty, we can conclude that this actual degree of the student's knowledge is located on the k-sigma interval $[\widetilde{P}-k\cdot\Sigma,\widetilde{P}+k\cdot\Sigma]$, where k depends on the desired degree of certainty: for $k=3$, we get the degree of certainty 99.9%, for $k=6$, we get the degree of certainty $1-10^{-8}$, etc.

This degree of knowledge could only increase, so we can conclude that the degree of knowledge cannot be smaller than the value

$$\underline{P}=\widetilde{P}-k\cdot\Sigma=\widetilde{P}-k\cdot\sqrt{\frac{\widetilde{P}\cdot(1-\widetilde{P})}{S}}. \quad (9)$$

This leads is to the following recommendation.

How to take into account intermediate grades: recommendation. Let us assume that out of s sub-questions on the final exam, the student answered n sub-questions correctly. Let us also assumed that out of S sub-questions asked before the final exam, the student answered N sub-questions correctly. Then, the actual student's degree of knowledge p can be described by a normal distribution with mean

$$\widetilde{p}=\frac{n}{s} \quad (10)$$

and standard deviation

$$\sigma\approx\sqrt{\frac{\widetilde{p}\cdot(1-\widetilde{p})}{s}} \quad (11)$$

restricted to values

$$p\geq\widetilde{P}-k\cdot\sqrt{\frac{\widetilde{P}\cdot(1-\widetilde{P})}{S}}, \quad (12)$$

where we denoted

$$\widetilde{P}=\frac{N}{S} \quad (13)$$

Thus, a natural measure of the student's knowledge is the mean value of this restricted normal distribution. The larger the intermediate grade $\widetilde{P}$, the larger the restricting lower bound on p and thus, the larger the resulting mean. So, we indeed take into account the intermediate grades when estimating the overall grade for the class.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes).

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.

References

1. D. J. Sheskin, *Handbook of Parametric and Non-Parametric Statistical Procedures*, Chapman & Hall/CRC, London, UK, 2011.