Why Grade Distribution Is Often Multi-Modal: an Uncertainty-Based Explanation

Olga Kosheleva, Christian Servin, and Vladik Kreinovich

Abstract There are many different independent factors that affect student grades. There are many physical situations like this, in which many different independent factors affect a phenomenon, and in most such situations, we encounter normal distribution – in full accordance with the Central Limit Theorem, which explains that in such situations, distribution should be close to normal. However, the grade distribution is definitely not normal – it is multi-modal. In this paper, we explain this strange phenomenon, and, moreover, we explain several observed features of this multi-modal distribution.

1 Formulation of the Problem

Many different factors affect the student grades. Many different independent factors affect the student's grade in a class. The grade can be affected by a student's preparedness for different sections of the material, by the student's degree of involvement in other classes, by how well the professor's teaching style matches the student's learning style, by possible personal problems – the list can go on and on.

Based on this, one would expect normal distribution of the grades. Situations when the result comes from the joint effect of a large number of independent factors are ubiquitous in real life. From the mathematical viewpoint, such situations are well

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

Christian Servin
Computer Science and Information Technology Systems Department
El Paso Community College, 919 Hunter, El Paso, TX 79915, USA, e-mail: cservin@gmail.com

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

analyzed. It is known that the distribution of the sum of a large number of relatively small independent random variables is close to Gaussian (normal) — this is the gist of the so-called Central Limit Theorem; see, e.g., [16]. And indeed, normal distributions are encountered in many such situations.

Based on the the above explanation, one would expect that in a large class, grades would also be normally distributed. But they are not.

A puzzling fact: grade distribution is multi-modal. Even for a relatively large class, we very rarely see the bell-shaped curve of a normal distribution. In reality, the distribution is multi-modal.

This multi-modality is a well-known phenomenon: so well-known that many professors use it for grading. To avoid a natural student's complaint about grading fairness, that this student with 89.9 got a B but someone with a practically indistinguishable grade of 90.1 gets an A, experienced teachers recommend to use, as an A-or-B threshold, to select not some arbitrary number like 90, but the largest gap between the grades which is close to 90. This way, there is a significant gap between B and A students, and thus, the grades are viewed as more fair than before.

Such a gap can always be found – exactly because the distribution is multi-modal, because between the modes, the probability density gets very low, and thus, gaps between neighboring grades become much larger than in the vicinity of each mode.

But why? It is great that we can use multi-modality, but the question remains: why? In this paper, we provide a possible uncertainty-based explanation for this unexpected phenomenon.

Additional observations. Another interesting phenomenon is that the number of modes does not stay the same throughout the students' studies: for undergraduate students, we have more modes, whole for graduate students, we observe fewer modes. This seems to be in perfect accordance with the fact that in undergraduate studies, we usually use more different grades: A, B, and C, while for graduate students, C is practically a failure grade, so, in effect, we only use As and Bs. Convenient, but why?

Yet another convenient-but-why observation is that modes are almost equidistant, so the corresponding clusters are indeed close to the usual groupings of 90-100, 80-90, 70-80, etc.

In this paper, we try to explain these additional observations as well.

2 Analysis of the Problem and the Resulting Qualitative Explanation

Important phenomenon: students help each other. At first glance, the situation with grades is the same as with other cases when Central Limit Theorem works: e.g., in situations like Brownian motion where the motion of a particle is caused by a joint effect of many different phenomena.

At first glance, the situation is the same, and if the students were simply randomly affected by all the factors mentioned above, we probably would have observed exactly the same normal distribution as in many physical situations.

But there is a big difference between students and particles: students help each other. This help may not always be a huge contribution to the student's success, but, as everyone who has ever studied knows well, it does provide an important help. Students ask questions to each other, students exchange ideas – and often even form study groups to study together, and it helps.

How does this helping phenomenon affects the resulting grade distribution?

What happens when two students study together: ideal case. In the ideal case, when two students study with each other, they exchange knowledge, and at the end, both get the exact same amount of knowledge. To be more precise, each student knows exactly what he knew before + what the other student knew. As a result, if at this moment, we give them a test, they will get the exact same grade – reflecting their exact same state of knowledge.

What happens in practice. In practice, this ideal exchange of information only happens when students are at approximately the same level of knowledge. If we try to bring together two students with a big gap between them – e.g., a straight A student and an almost-failing student – this rarely helps, because most students lack the ability to clearly explain things to those who know much less.

For example, in VK's department, when we started hiring undergraduate instructional assistants for classes, it turns out that students who in their time got B for the corresponding classes were much better in helping new students than those who got A – the A students knew material much better, but they could not as convincingly explain it to the new students.

Not only students, starting professors (and even some should-have-been-experienced professor) have the same limitation.

As a result, this exchange of knowledge happens only when the difference between the students' levels of knowledge – i.e., the difference between their grades – is small. The larger the difference, the less probable it is that the knowledge exchange will happen.

Eventually, the students get better in this knowledge exchange skills: as they progress from undergraduate students to graduate ones, their ability improves, and the threshold beyond which they cannot effectively exchange knowledge decreases.

What happens as a result: a qualitative description. How does the existence of this collaboration gap affect distribution of student grades? To understand the effect on the qualitative level, let us consider a simplified model of student grade distribution.

Suppose that originally, students' knowledge levels are uniformly distributed – at least on some segment of the grades interval. This can be simplified into saying that the students' grades are initially distributed with the same step h. In other words, these grades, when sorted in the increasing order, form the following sequence:

$$g_0 < g_1 = g_0 + h < g_2 = g_0 + 2h < \ldots < g_k = g_0 + k \cdot h < \ldots < g_n = g_0 + n \cdot h.$$

In the beginning, the least well performing student is the one who is the most desperate for help. So, it is reasonable to expect that first, the student whose grade is g_0 will reach for help. The person most appropriate for helping him is the person whose grade is the closest to his – e.g., the student with the grade g_1. They start actively collaborating, and, as the result of this collaboration, the reach the same grade level – which is close to g_1. For simplicity, let us assume that their grade level is now exactly g_1.

The student whose grade is g_2 also needs help — so he/she contacts the closest better student, the one with the grade level g_3. As a result of their collaboration, they both reach the same level g_3.

Similarly, the first yet-unpaired student g_4 teams us with g_5, so their grade level is now g_5, etc. As a result of this first round of exchanges, we have pairs of students whose grades are

$$g_1 = g_0 + h, g_3 = g_0 + 3h, g_5 = g_0 + 5h, \ldots$$

note that the gap between different levels has doubled, from h to $2h$.

Now, the same process starts again: students at level g_1 are the most eager for help, so they contact students at the next level g_3 to form a study group. As a result of their joint study, all four of them reach the level g_3.

Students at the lowest not-yet-involved level g_5 contact students from the level g_7 and all get to the level g_7, etc. Now, we have a new list of grades:

$$g_3 = g_0 + 3h < g_7 + 7h < g_{11} = g_0 + 11h < \ldots$$

The gap has doubled again, to $4h$. At the next iteration of this process, the gap will double again – until it reaches the threshold after which the mutual exchange of knowledge becomes difficult.

As a result, instead of the original uniform distribution, we have big groups with approximately the same level of knowledge – separated by gaps in which there are no students with this particular grade.

From simplified model to real life situations. Of course, the above description is oversimplified. In reality, the original distances $g_i - g_{i-1}$ are not exactly equal, and the effects are also not always the same. As a result, what we get is not the above simplified picture, but rather a smoothed version of it: instead of groups of students with identical grades, we have groups with close grades – i.e., in effect, we will have a multi-modal distribution.

So, the mutual help indeed explains why grade distribution is multi-modal.

Why there are fewer modes for grades of graduate students? The same phenomenon explains why for graduate students, we usually have fewer modes than for undergraduate ones: graduate students have already learned how to exchange knowledge, so for them, the threshold above which they cannot productively collaborate is much higher. As a result, they continue merging into a single cluster even when

at the undergraduate level, we would have reached the original merging threshold and stopped. As a result, for graduate students, we have fewer clusters – i.e., fewer modes of the resulting distribution.

3 Towards a Quantitative Analysis

Analogy with physics. To looks for a quantitative analysis of the situation, let us look for other situations when a similar phenomenon occurs.

What is the above phenomenon? We started with a distribution which was perfectly uniform. This distribution was symmetric, in the sense that it does not change – at least locally – if we simply shift all the grades by the same number h. Then what happens if two nearby students with grades g_i and g_{i+1} start collaborating. As a result, the knowledge of both students reaches the same level g_{i+1}: $g'_i = g'_{i+1} = g_{i+1}$. Now, we get a gap of width $2h$ between the levels $g'_{i-1} = g_{i-1}$ and $g'_i = g_{i+1}$.

The distribution is no longer invariant with respect to a shift by h – even if many pairs exchange their knowledge. The original symmetry is broken.

This phenomenon of spontaneous symmetry breaking is ubiquitous in physics; see, e.g., [2, 17]. We can easily observe this phenomenon: e.g., if we drop a breakable rotationally symmetric vase, it will not break into rotationally symmetric pieces: it will break into irregular ones.

This phenomenon is very important: without it, our Universe would remain the same highly homogenous and isotropic blurb that it was close to the Big Bang. Luckily, gravity acts as the spontaneous symmetry breaking mechanism. Specifically, if a small fluctuation appears and at some location, the density at this location becomes slightly larger than at other locations, then this heaver location will start attracting other particles. As it attracts them, its mass increases and it attracts more and more – until the whole original homogeneous cloud disintegrates into what we call proto-galaxies [3, 4, 8].

In physics, researchers go beyond qualitative explanations. Not only this mechanism explains symmetry breaking, it explains all the observed shapes of celestial bodies, such as spiral galaxies and planetary systems like ours in which distances of the planets to the central star form a geometric progression; see, e.g., [3, 4, 8]. (This mechanics also explains relative frequencies of different shapes.)

To understand the corresponding explanations, we need to know the basic ideas of statistical physics, according to which it is not very probable to go from a completely symmetric state to a state with no symmetries at all: it is much more probable that – at least at first – some symmetries will be preserved, and the more symmetries will be preserved, the more probable the corresponding transition. For example, a solid body (i.e., matter in highly symmetric – usually crystal – state), when heated, usually does not immediately gets transformed into a completely asymmetric state of gas, it first gets transformed into the state of the liquid in which some symmetries are preserved [2, 17].

What are the symmetries in the grades case? How can we apply the above idea in our example – of grade distribution?

In the gravity case, the original symmetries were easy to find: rotations, shifts, probably scalings. To apply a similar approach to grade distribution, we need to understand what are the natural symmetries here. Let us brainstorm.

How are grades formed? Usually, by simply adding the grades corresponding to different assignments – and these grades, in their turn, are obtained by simply adding grades on different problems or parts or aspects of each assignment. Some assignments are very tough, some are much easier. There have to be easier assignments: we are talking mass education, not training students to win at an international student olympiad in computer science.

What does it mean that the assignment is relatively easy? That on this particular assignment (or part of the assignment), practically all the students will get a very good grade. One professor may give a certain number of such assignments, another professor may give one more such relatively simple task. The difference between the grades given by these two professors will be exactly the grade e on this extra assignment.

So, depending on who teaches a class, for the same level of knowledge, students may get grades g_i from one professor and grades $g_i + e$ from another one. This shift $g_i \rightarrow g_i + e$ is therefore a reasonable symmetry here. In other words, the original situation is invariant under all possible shifts $g \rightarrow g + e$.

Now, spontaneous symmetry breaking occurs, and the situation is no longer fully symmetric. However, in line with the general ideas from statistical physics, the most probable situation is that *some* of the original symmetries will remain. In other words, there remains some value e_0 so that – at least locally – the resulting distribution will not change if we simply add e_0 to all the grades. In particular, this means that if we add e_0 to one mode (i.e., to one local maximum of the corresponding probability distribution), then we should again encounter a model — i.e., yet another local maximum. So, in the first approximation, local maxima (modes) are almost equidistant – and, as we have mentioned, this is exactly what we observe!

Thus, this equidistance distribution can also be explained by our analysis.

Future work. To make the conclusions more qualitative, we need to provide a formal explanation of the threshold. In the above analysis, we viewed the threshold as, in effect, an interval beyond which collaboration is not productive – in line with the interval uncertainty (see, e.g., [5, 7, 9, 11, 12]). However, in practice, this threshold is not precise, it is imprecise – so we believe that the use of fuzzy techniques (see, e.g., [1, 6, 10, 13, 14, 15, 18]) will lead to an even better description of this phenomenon.

Acknowledgments

References

1. R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.
2. R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.
3. A. Finkelstein, O. Kosheleva, and V. Kreinovich, "Astrogeometry: towards mathematical foundations", International *Journal of Theoretical Physics*, 1997, Vol. 36, No. 4, pp. 1009–1020.
4. A. Finkelstein, O. Kosheleva, and V. Kreinovich, "Astrogeometry: geometry explains shapes of celestial bodies", *Geombinatorics*, 1997, Vol. VI, No. 4, pp. 125–139.
5. L. Jaulin, M. Kiefer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*, Springer, London, 2001.
6. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
7. V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing and Interval Computations*, Kluwer, Dordrecht, 1998.
8. S. Li, Y. Ogura, and V. Kreinovich, *Limit Theorems and Applications of Set Valued and Fuzzy Valued Random Variables*, Kluwer Academic Publishers, Dordrecht, 2002.
9. G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.
10. J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
11. R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.
12. S. G. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer, New York, 2005.
13. H. T. Nguyen and V. Kreinovich, "Nested intervals and sets: concepts, relations to fuzzy sets, and applications", In: R. B. Kearfott and V. Kreinovich (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996, pp. 245–290.
14. H. T. Nguyen, C. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.
15. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
16. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.
17. K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2017.
18. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.