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

Olga Kosheleva[1], Christian Servin[2], and Vladik Kreinovich[3]
Departments of [1]Teacher Education and [3]Computer Science
University of Texas at El Paso, El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu
[3]Computer Science and Information Technology Systems Department
El Paso Community College, 919 Hunter
El Paso, TX 79915, USA, cservin@gmail.com

1. 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.

2. 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.
- From the mathematical viewpoint, such situations are well analyzed.
- The distribution of the sum of many small independent random variables is close to Gaussian (normal).
- This is the gist of the so-called Central Limit Theorem.
- And indeed, normal distributions are encountered in many such situations.
- So one would expect that in a large class, grades would also be normally distributed.
- But they are not.

3. 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.
- There are natural student's complaints about grading fairness, when:
  - this student with 89.9 got a B but
  - someone with a practically indistinguishable grade of 90.1 gets an A.

4. Grade Distribution Is Multi-Modal (cont-d)

- 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.
- Thus, the grades are viewed as more fair than before.
- Such a gap can always be found – exactly because the distribution is multi-modal.
- Between the modes, the probability density gets very low.
- Thus, gaps between neighboring grades become much larger than in the vicinity of each mode.

5. But Why?

- It is great that we can use multi-modality, but the question remains: why?
- In this talk, we provide a possible uncertainty-based explanation for this unexpected phenomenon.
- Another interesting phenomenon is that the number of modes does not stay the same:
  - for undergraduate students, we have more modes,
  - while 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.

6. But Why (cont-d)

- 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 talk, we try to explain these additional observations as well.

7. 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 Brownian motion, the motion of a particle is caused by a joint effect of many different phenomena.
- At first glance, the situation is the same:
  - 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.

8. Students Help Each Other (cont-d)

- This help may not always be a huge contribution to the student's success.
- However, as everyone who has ever studied knows well, it does provide an important help.
- Students ask questions to each other, exchange ideas, study together, and it helps.
- How does this helping phenomenon affect the resulting grade distribution?

9. What Happens When Two Students Study Together: Ideal Case

- In the ideal case, when two students study with each other, they exchange knowledge.
- 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.

10. 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, this rarely helps.
- Indeed, 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 were much better in helping new students than those who got A.

11. What Happens in Practice (cont-d)

- "A" students knew material better, but they could not as convincingly explain it to the students.
- This exchange of knowledge happens only when 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.
- The threshold beyond which they cannot effectively exchange knowledge decreases.

12. 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 grade distribution.
- Suppose that originally, students' knowledge levels are uniformly distributed – at least on some 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:

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

13. What Happens as a Result (cont-d)

- 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: grade g_1.
- They start actively collaborating, and reach the same grade level – 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.

14. What Happens as a Result (cont-d)

- 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.

15. What Happens as a Result (cont-d)

- 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.
- 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.
- Eventually, it reaches the threshold after which the mutual exchange of knowledge becomes difficult.

16. What Happens as a Result (cont-d)

- 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.

17. 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.
- In effect, we will have a multi-modal distribution.
- So, the mutual help indeed explains why grade distribution is multi-modal.

18. Why There Are Fewer Modes for Grades of Graduate Students?

- The same phenomenon explains why for graduate students, we usually have fewer modes.
- 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.

19. Analogy with Physics

- To look for a quantitative analysis of the situation, let us look for other similar situations.
- 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}: $g'_i = g'_{i+1} = g_{i+1}$.

20. Analogy with Physics (cont-d)

- 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 w.r.t. 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.
- 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 homogenous and isotropic blurb as in the Big Bang.

21. Analogy with Physics (cont-d)

- 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,
  - then this heavier location will start attracting other particles.
- As it attracts them, its mass increases and it attracts more and more.
- In the end, the whole original homogeneous cloud disintegrates into what we call proto-galaxies.

22. 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 to the star form a geometric progression.
- This mechanics also explains relative frequencies of different shapes.
- To understand the corresponding explanations, we need to know the basic ideas of statistical physics.

23. In Physics (cont-d)

- According to statistical physics, 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.
- 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.

24. 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 symmetries here.
- Let us brainstorm. How are grades formed?
- Usually, by simply adding the grades corresponding to different assignments.
- These grades, in their turn, are obtained by simply adding grades on different parts of each assignment.
- Some assignments are very tough, some are much easier.
- There have to be easier assignments.

25. What Are the Symmetries (cont-d)

- Indeed, we are talking mass education, not training students to win at an international student olympiad.
- What does it mean that the assignment is relatively easy?
- That on this particular 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 is the grade e on this extra assignment.

26. What Are the Symmetries (cont-d)

- 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 $g_i \to g_i + e$ is therefore a reasonable symmetry here.
- In other words, the original situation is invariant under all possible shifts $g \to g + e$.
- 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.

27. What Are the Symmetries (cont-d)

- 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 mode, i.e., yet another local maximum.
- So, in the first approximation, local maxima (modes) are almost equidistant.
- This is exactly what we observe!
- Thus, this equidistance distribution can also be explained by our analysis.

28. Future Work

- To make the conclusions more qualitative, we need to provide a formal explanation of the threshold.
- We viewed the threshold as, in effect, an interval beyond which collaboration is not productive.
- This is in line with the interval uncertainty.
- However, in practice, this threshold is not precise, it is imprecise.
- So we believe that the use of fuzzy techniques will lead to an even better description of this phenomenon.

29. Acknowledgments

This work was supported in part by the US National Science Foundation via grant HRD-1242122 (Cyber-ShARE).