How to Explain Empirical Distribution of Software Defects by Severity

Francisco Zapata[1], Olga Kosheleva[2], and Vladik Kreinovich[3]

Departments of [1]Industrial, Manufacturing, and Systems Engineering
[2]Teacher Education, and [3]Computer Science
University of Texas at El Paso, El Paso, TX 79968, USA
fazg74@gmail.com, olgak@utep.edu, vladik@utep.edu

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

1. Classification of Software Defects

- Software packages have defects of different severity.
- Some defects allow hackers to enter the system and can thus, have a potentially high severity.
- Other defects are minor and maybe not be worth the effort needed to correct them.
- For example:
  - if we declare a variable which is never used
  - or we declare an array of too big size, so that most of its elements are never used,
  - this makes the program not perfect,
  - but its only negative consequences are wasting computer time or memory.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

2. Classification of Software Defects (cont-d)

- In the last decades, several tools have appeared that:
  - given a software package,
  - mark possible defects of different potential severity.
- Usually, software defects which are worth repairing are classified into three categories by their relative severity:
  - software defects of very high severity (they are also known as *critical*);
  - software defects of high severity (they are also known as *major*); and
  - software defects of medium severity.
- This is equivalent to classifying *all* the defects into four categories – the fourth category are *minor* defects.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

3. Cautious Approach

- The main objective of this classification is not to miss any potentially serious defects.
- Thus, in case of any doubt, a defect is classified into the most severe category possible.
- As a result:
  - the only time when a defect is classified into medium severity category
  - is when we are absolutely sure that this defect is not of high or of very high severity.
- If we have any doubt, we classify this defect as being of high or very high severity.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

4. Cautious Approach (cont-d)

- Similarly:
  - the only time when a defect is classified as being of high severity
  - is when we are absolutely sure that this defect is not of very high severity.
- If there is any doubt, we classify this defect as being of very high severity.
- In particular:
  - in situations in which we have no information about severity of different defects,
  - we should classify them as of very high severity.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

5. Cautious Approach (cont-d)

- As we gain more information about the consequences of different defects, we can start
  - assigning some of the discovered defects
  - to medium or high severity categories.
- However, since by default we classify a defect as having high severity:
  - the number of defects classified as being of very high severity should still be the largest,
  - followed by the number of defects classified as being of high severity,
  - and finally, the number of defects classified as being of medium severity should be the smallest.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

6. Empirical Results

- Software defects can lead to catastrophic consequences.
- So, it is desirable to learn as much as possible about different defects.
- In particular, it is desirable to know how frequently one meets defects of different severity.
- For this, it is necessary to study a sufficiently large number of detects.
- This is a challenging task:
  - while many large software packages turn out to have defects, even severe defects,
  - such defects are rare, uncovering each severe defect in a commercial software is a major event.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

7. Empirical Results (cont-d)

- At first glance, it may seem that:
  - unless we consider nor-yet-released still-being tested software,
  - there is no way to find a sufficient number of defects to make statistical analysis possible.
- However, there is a solution to this problem: namely, legacy software.
- This software that was written many years ago, when the current defect-marking tools were not yet available.
- Legacy software works just fine:
  - if it had obvious defects,
  - they would have been noticed during its many years of use.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

8. Empirical Results (cont-d)

- However, legacy software works just fine only in the original computational environment.
- When the environment changes, many hidden severe defects are revealed.
- In some cases, the software package used a limited size buffer that was sufficient for the original usage.
- When the number of users increases, the buffer becomes insufficient.
- Another typical situation is when a C program, which should work for all computers,
  - had some initial bugs that were repaired by hacks
  - that explicitly took into account that in those days, most machines used 32-bit words.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

9. Empirical Results (cont-d)

- As a result, when we run the supposedly correct program on a 64-bit machine, we get many errors.
- Such occurrences are frequent.
- As a result, when hardware is upgraded – e.g., from 32-bit to 64-bit machines,
  - software companies routinely apply the defect-detecting software packages to their legacy code
  - to find and repair all the defects of very high, high, and medium severity.
- Such defects are not that frequent; however
  - even if in the original million-lines-of codes software package, 99.9% of the lines of code are flawless,
  - this still means that there may be a thousand of severe defects.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

10. Empirical Results (cont-d)

- This is clearly more than enough to provide a valid statistical analysis of distribution of software defects by severity.
- So, at first glance, we have a source of defects.
- However, the problem is that companies – naturally – do not like to brag about defects in their code.
- This is especially true when a legacy software turns out to have thousands of severe defects.
- On the other hand, companies are interested in knowing the distribution of the defects:
  - this would help them deal with these defects,
  - and not all software companies have research department that would undertake such an analysis.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

11. Empirical Results (cont-d)

- In view of this interest, we contacted software folks from several companies.
- They allowed us to test their legacy code and record the results – on the condition that
  - when we publish the results,
  - we would not disclose the company names.
- We are not even allowed
  - to disclose which defect-detecting tool was used
  - because many companies use their own versions of such tools.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

12. Empirical Results (cont-d)

- Interestingly, for different legacy software packages:
  - as long as they were sufficiently large (and thus containing a sufficiently large number of defects),
  - the distribution of defects by severity was approximately the same.
- In this talk, we illustrate this distribution on three typical cases; see the Table.

	Case 1	Case 2	Case 3
Total number of defects	996	1421	1847
Very high severity defects	543	738	1000
High severity defects	320	473	653
Medium severity defects	133	210	244

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

13. Analysis of the Empirical Results: General Case

- In all 3 cases, the numbers of very high, high, and medium severity defects follow the ratio $5:3:1$.
- In other words:
  - the proportion of software defects of very high severity is close to $\dfrac{5}{5+3+1} = \dfrac{5}{9} \approx 56\%$;
  - the proportion of software defects of high severity is close to $\dfrac{3}{5+3+1} = \dfrac{3}{9} \approx 33\%$; and
  - the proportion of software defects of medium severity is close to $\dfrac{1}{5+3+1} = \dfrac{1}{9} \approx 11\%$.
- Let us show it on the example of the above three cases; the match is up to $\approx 20\%$ accuracy.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

Case 1	actual #	predicted #
Very high severity defects	543	553
High severity defects	320	332
Medium severity defects	133	111

Case 2	actual #	predicted #
Very high severity defects	738	789
High severity defects	473	474
Medium severity defects	210	158

Case 3	actual #	predicted #
Very high severity defects	1000	1026
High severity defects	653	616
Medium severity defects	244	206

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

14. How to Explain the Empirical Distribution

- We want to find the three frequencies:
  - the frequency p_1 of defects of medium severity;
  - the frequency p_2 of defects of high severity, and
  - the frequency p_3 of defects of very high severity.
- All we know is that $p_1 < p_2 < p_3$.
- We will use ideas related to interval uncertainty to explain the above empirical dependence.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

15. First Idea: Let Us Use Intervals Instead of Exact Numbers

- As we have seen, the frequencies somewhat change from one example to another.
- So, instead of selecting single values p_1, p_2, and p_3, we should select three *regions* of possible values.
- So, we should select:
  - an interval $[\underline{F}_1, \overline{F}_1]$ of possible values of p_1;
  - an interval $[\underline{F}_2, \overline{F}_2]$ of possible values of p_2; and
  - an interval $[\underline{F}_3, \overline{F}_3]$ of possible values of p_3.
- To guarantee that $p_1 < p_2$, we want to make sure that:
  - every value from the first interval $[\underline{F}_1, \overline{F}_1]$
  - is smaller than or equal to any value from the second interval $[\underline{F}_2, \overline{F}_2]$.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

16. First Idea (cont-d)

- To guarantee this, it is sufficient to require that:
  - the largest value $\overline{F}_1$ from the first interval
  - is smaller than or equal to the smallest value of the second interval: $\overline{F}_1 \le \underline{F}_2$.
- Similarly, to guarantee that $p_2 < p_3$, we want to make sure that
  - every value from the second interval $[\underline{F}_2, \overline{F}_2]$
  - is smaller than or equal to any value from the third interval $[\underline{F}_3, \overline{F}_3]$.
- To guarantee this, it is sufficient to require that
  - the largest value $\overline{F}_2$ from the first interval
  - is smaller than or equal to the smallest value of the third interval: $\overline{F}_2 \le \underline{F}_3$.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

17. First Idea Expanded: Let Us Make These Intervals As Wide As Possible

- We decided to have intervals of possible values of p_i instead of exact values of the frequencies.
- To fully follow this idea, let us make these intervals as wide as possible.
- Let us make sure that it is not possible to increase one of the intervals without violating the inequalities.
- This means that we should have no space left between $\overline{F}_1$ and $\underline{F}_2$.
- Otherwise, we can expand either the first or the second interval.
- We should therefore have $\overline{F}_1 = \underline{F}_2$.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

18. First Idea Expanded (cont-d)

- Similarly, we should have no space between $\overline{F}_2$ and $\underline{F}_3$.
- Otherwise, we can expand either the second or the third interval.
- We should therefore have $\overline{F}_2 = \underline{F}_3$.
- Also, we should have $\underline{F}_1 = 0$ – otherwise, we can expand the first interval.
- As a result, we get the division of the interval $[0, F]$ of possible frequencies into three sub-intervals:
  - the interval $\left[0, \overline{F}_1\right]$ of possible values of p_1;
  - the interval $\left[\overline{F}_1, \overline{F}_2\right]$ of possible values of p_2; and
  - the interval $\left[\overline{F}_2, F\right]$ of possible values of p_3.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

19. Second Idea: We Have No Reason to Take Intervals of Different Widths

- We have no a priori reason to assume that the three intervals have different widths.
- Thus, it is reasonable to assume that these three intervals have the exact same width, i.e., that

$$\overline{F}_1 = \overline{F}_2 - \overline{F}_1 = F - \overline{F}_2.$$

- From $\overline{F}_2 - \underline{F}_1 = \overline{F}_1$, we conclude that $\overline{F}_2 = 2\overline{F}_1$.
- Now, from the condition that $F - \overline{F}_2 = \overline{F}_1$, we conclude that $F = \overline{F}_2 + \overline{F}_1 = 2\overline{F}_1 + \overline{F}_1 = 3\overline{F}_1$.
- So, we have the following three intervals:
  - the interval $\left[0, \overline{F}_1\right]$ of possible values of p_1;
  - the interval $\left[\overline{F}_1, 2\overline{F}_1\right]$ of possible values of p_2; and
  - the interval $\left[2\overline{F}_1, 3\overline{F}_1\right]$ of possible values of p_3.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

20. Third Idea: Which Value From the Interval Should We Choose

- We would like to select a single "typical" value from each of the three intervals.
- If we know the probability of different values from each interval, we could select the average value.
- We do not know these probabilities.
- So to use this approach, we need to select one reasonable probability distribution on each interval.
- A priori, we have no reason to believe that some values from a given interval are more probable than others.
- Thus, it is reasonable to conclude that all the values within each interval are equally probable.
- So, on each of the three intervals, we have a uniform distribution.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

21. Now, We Are Ready to Produce the Desired Probabilities

- For the uniform distribution on an interval, the mean value is the midpoint of the interval.
- So, as the estimate for p_1, we select the midpoint of the first interval $\left[0, \overline{F}_1\right]$: $p_1 = \dfrac{0 + \overline{F}_1}{2} = \dfrac{\overline{F}_1}{2}$.
- As the estimate for p_2, we select the midpoint of the second interval $\left[\overline{F}_1, 2\overline{F}_1\right]$: $p_2 = \dfrac{\overline{F}_1 + 2\overline{F}_1}{2} = 3 \cdot \dfrac{\overline{F}_1}{2}$.
- As the estimate for p_1, we select the midpoint of the third interval $\left[2\overline{F}_1, 3\overline{F}_1\right]$: $p_3 = \dfrac{2\overline{F}_1 + 3\overline{F}_1}{2} = 5 \cdot \dfrac{\overline{F}_1}{2}$.
- We see that $p_2 = 3p_1$ and $p_3 = 5p_1$, i.e., $1 : 3 : 5$.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

22. What If We Had a More Detailed Classification – Into More Than Three Severity Levels?

- In the above analysis, we used the usual subdivision of all non-minor software defects into three levels:
  - of medium severity,
  - of high severity, and
  - of very high severity.
- However:
  - while such a division is most commonly used,
  - some researchers and practitioners have proposed a more detailed classification,
  - when, e.g., each of the above three levels is further subdivided into sub-categories.
- To the best of our knowledge, such detailed classification have not yet been massively used in industry.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

23. What If We Had a More Detailed Classification (cont-d)

- Thus, we do not have enough data to find the empirical distribution of software defects by sub-categories.
- However:
  - while we do not have the corresponding empirical data,
  - we can apply our theoretical analysis to come up with reasonable values of expected frequencies.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

24. Analysis of the Problem

- We consider the situation when each of the three severity levels is divided into several sub-categories.
- Let us denote the number of sub-categories in a level by k.
- Then, overall, we have $3k$ sub-categories of different level of severity.
- Similar to the previous section, it is reasonable to conclude that
  - each of these categories
  - corresponds to an interval of possible probability values:

$$\left[0, \overline{F}_1\right], \left[\overline{F}_1, \overline{F}_2\right], \left[\overline{F}_2, \overline{F}_3\right], \ldots, \left[\overline{F}_{3k-1}, \overline{F}_{3k}\right].$$

- Also, similarly to the previous section, it is reasonable to require that all these intervals are of the same length.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

25. Analysis of the Problem (cont-d)

- Thus, the intervals have the form

$$\left[0, \overline{F}_1\right], \left[\overline{F}_1, 2\overline{F}_1\right], \left[2\overline{F}_1, 3\overline{F}_1\right], \ldots, \left[(3k-1)\cdot\overline{F}_1, (3k)\cdot\overline{F}_1\right].$$

- As estimates for the frequencies, we take midpoints of the intervals:

$$\frac{\overline{F}_1}{2}, \quad 3\cdot\frac{\overline{F}_1}{2}, \quad 5\cdot\frac{\overline{F}_1}{2}, \ldots, \quad \frac{6k-1}{2}\cdot\frac{\overline{F}_1}{2}.$$

- This analysis shows that the frequencies of detects of different levels of severity follow the ratio

$$1 : 3 : 5 : \ldots : (6k-1).$$

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

26. This Result Is Consistent With Our Previous Findings

- For each of three main severity levels, we can combine all the defects from different sub-categories of this level.
- Will we still get the same ratio $1:3:5$ as before?
- The answer is "yes"; indeed:
  - by adding up k sub-categories with lowest severity,
  - we can calculate the total frequency of medium severity defects as

$$(1+3+5+\ldots+(2k-5)+(2k-3)+(2k-1))\cdot\frac{\overline{F}_1}{2}.$$

- In the above sum of k terms:
  - adding the first and the last terms leads to $2k$;
  - similarly, adding the second and the last but one terms lead to $2k$, etc.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

27. This Result Is Consistent With Our Previous Findings (cont-d)

- For each pair, we get $2k$.
- Out of k terms, we have $\dfrac{k}{2}$ pairs, so the overall sum is

$$1+3+5+\ldots+(2k-5)+(2k-3)+(2k-1) = \frac{k}{2}\cdot(2k) = k^2.$$

- Thus, the total frequency of medium severity defects is $k^2 \cdot \dfrac{\overline{F}_1}{2}$.
- Similar, the total frequency of high severity defects is equal to

$$((2k+1)+(2k+3)+\ldots+(4k-3)+(4k-1))\cdot\frac{\overline{F}_1}{2}.$$

- Here, similarly, the sum is

$$(2k+1)+(2k+3)+\ldots+(4k-3)+(4k-1) = \frac{k}{2}\cdot(6k) = 3k^2.$$

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

28. This Result Is Consistent With Our Previous Findings (cont-d)

- So, the total frequency of high severity defects is

$$3k^2 \cdot \frac{\overline{F}_1}{2}.$$

- The total frequency of very high severity defects is

$$((4k+1)+(4k+3)+\ldots+(6k-3)+(6k-1)) \cdot \frac{\overline{F}_1}{2}.$$

- Here similarly, the sum is equal to

$$(4k+1)+(4k+3)+\ldots+(6k-3)+(6k-1) = \frac{k}{2}\cdot(10k) = 5k^2.$$

- So, the total frequency of high severity defects is

$$5k^2 \cdot \frac{\overline{F}_1}{2}.$$

- These frequencies indeed follow the ratio $1:3:5$.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

29. What If We Have 4 or More Original Severity Levels?

- Another possible alternative to the usual 3-level scheme is to have not 3, but 4 or more severity levels.
- In this case,
  - if we have L levels,
  - then a similar analysis leads to the conclusion that the frequencies should follow the ratio

$$1 : 3 : 5 : \ldots : (2L - 1).$$

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .

30. Acknowledgments

This work was supported in part by the US National Science Foundation grant HRD-1242122.

Classification of . . .
Cautious Approach
Empirical Results
How to Explain the . . .
First Idea: Let Us Use . . .
Second Idea: We Have . . .
Third Idea: Which . . .
What If We Had a . . .
What If We Have 4 or . . .