SIM Global Education

UOW AUSTRALIA

CSCI262 – SYSTEM SECURITY

Puzzle

14 October 2022

Standard Deviation, Variance and Mean

To compute the standard deviation for the distribution of the number of hashes needed, you need to use the standard deviation formula. Both population or sample standard deviation can be used.

I will explain based on population standard deviation because that’s what specified in the assignment question.

Standard Deviation, Variance and Mean

Definition:

Standard deviation is a measure of how spread out numbers are. It is computed as the square root of the **variance**, that is,

$$Standard\ deviation = \sigma = \sqrt{Variance}$$

Variance is the average of the **square differences** from the **mean**.

Mean is the average of a set of numbers.

Standard Deviation, Variance and Mean

For example, I have 10 numbers; 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10.

The **mean** is the average of these 10 numbers, which is obtained as **(1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10) / 10 = 5.5**.

The **variance** is then equals

$$\frac{(5.5-1)^2+(5.5-2)^2+(5.5-3)^2+(5.5-4)^2+(5.5-5)^2+(5.5-6)^2+(5.5-7)^2+(5.5-8)^2+(5.5-9)^2+(5.5-10)^2}{10}$$

$$= \mathbf{8.25}$$

Standard Deviation, Variance and Mean

The **standard deviation** is then equals

$$Standard\ deviation = \sigma = \sqrt{Variance}$$

$$= \sqrt{8.25}$$

$$= 2.87$$

Puzzles and Sub-puzzles

In the puzzle problem, to determine the average number of hashes needed, we can compute the mean (average) of the number of hashes needed.

Now we look at the an example similar (hopefully) to the assignment problem ☺

For puzzle A, we have one sub-puzzle with k = 6 bits.

With 6 bits, the worst expected hashes needed = $m \times 2^k$; where m is the number of sub-puzzle.

Puzzles and Sub-puzzles

Hence we have,

Worst expected hashes = $1 \times 2^6 = 64$ hashes.

For 64 hashes, what is the average number of hashes (mean)?

Average number of hashes = $\frac{\sum_1^n n}{n}$; where $n = 64$.

This is based on the definition of mean described above; that is,

Puzzles and Sub-puzzles

$$Mean = \frac{1 + 2 + 3 + \cdots + n}{n}$$

$$= \frac{\sum_1^n n}{n}, where \left(1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}\right)$$

We can calculate the average number of hashes using this formula:

Average number of hashes $= \frac{\sum_{n=1}^{64} n}{64} = \frac{\frac{64(64+1)}{2}}{64} = \frac{2080}{64} =$ 32.5 hashes.

Puzzles and Sub-puzzles

- Next, we look at how to compute the standard deviation:

$$Standard\ deviation = \sigma = \sqrt{Variance}$$

and

Variance is the average of the square differences from the mean.

Hence,

$$Variance = \frac{(32.5 - 1)^2 + (32.5 - 2)^2 + \cdots + (32.5 - 64)^2}{64}$$

Puzzles and Sub-puzzles

- Square difference from mean = $(32.5 - 1)^2$, that is, the square of the difference of the number from the mean. We have computed the mean to be 32.5, and we have 64 difference values (number of hashes), thus we need to find the average of these 64 different terms.

- You can use a statistic calculator with variance function to compute it, otherwise, you need to do the long way of finding the difference of each value from the mean, square the result and add them all before divide the overall value by 64.

Puzzles and Sub-puzzles

- I used a MS Excel to calculate ☺, and I got the variance as follow:

  Variance = 341.25.

- The standard deviation is then the square of the variance; that is,

  Standard deviation = $\sigma = \sqrt{341.25} = 18.47$.

Puzzles and Sub-puzzles

- How about puzzle with multiple sub-puzzles?

- The calculation should be similar. We just need to multiply each working with the number of sub-puzzles. For example, for Puzzle B, there are two sub-puzzle with k = 5.

- First we need to compute the worst expected hash per sub-puzzle, that is, individual sub-puzzle = $1 \times 2^k = 1 \times 2^5 = 32$.

Puzzles and Sub-puzzles

- Average number of hashes per sub-puzzle = $\frac{\sum_{n=1}^{32} n}{32} = \frac{528}{32} = 16.5$

- Since there are two sub-puzzle in the Puzzle B, hence the average number of hashes for puzzle B = 16.5 x 2 = 33.

Puzzles and Sub-puzzles

- As for the standard deviation, it can be calculated as follows:

$$Variance = \frac{(16.5-1)^2 + (16.5-2)^2 + (16.5-3)^2 + \cdots + (16.5-32)^2}{32}$$

$$= 85.25$$

- Since there are two sub-puzzle, the total variance = 85.25 x 2 = 170.5.

- Hence the standard deviation = $\sqrt{170.5} = 13.05$

Puzzles and Sub-puzzles

- Okay now for your question related to the standard deviation from the lecture slide (SIM-2017-S2-CSCI262-S4b, slide 12), the standard deviation 2.2404 is obtained as follows:

- Four puzzle, with k = 2 each.
- Hence the worst expected hash per sub-puzzle = $1 \times 2^2 = 4$.
- Average number of hashes = $\frac{\sum_{n=1}^{4} n}{4} = \frac{10}{4} = 2.5\ per\ puzzle$.

- Since there are four sub-puzzles, hence the average number of hashes (expected hashes) = 2.5 x 4 = 10.

Puzzles and Sub-puzzles

The variance $= \frac{(2.5-1)^2+(2.5-2)^2+(2.5-3)^3+(2.5-4)^2}{4} = 1.25$

Since there are 4 sub-puzzle, the total variance = 5

Hence the standard deviation = $\sigma = \sqrt{5} \approx 2.24$.