Analysis of Student Activity Simulation with Simplified PFA Model

Sepehr Abbaspour

January 2025

1 Evaluation

1.1 How can you measure the goodness of fit to compare the different simulation settings and find out the simulation setting that is closest to the gold standard? Propose at least two different ways to evaluate the goodness of fit.

1. **Mean Absolute Error (MAE):** MAE calculates the average absolute difference between predicted values and actual values. It's easy to understand and interpret, making it a straightforward measure of error. It is on the same scale as the data, which means it's not influenced by the magnitude of the values and remains interpretable.

2. **Mean Squared Error (MSE):** MSE squares the errors, giving more weight to larger errors. This is useful when we want to penalize larger deviations more heavily, as it can highlight models that make large mistakes. It has desirable mathematical properties, making it easier to work with in optimization problems. For example, it's differentiable and its gradient is straightforward to compute, which is beneficial in optimization algorithms like gradient descent. Because MSE amplifies larger errors, it can better differentiate between models when errors are distributed unevenly. It highlights models that perform better overall by minimizing large errors more effectively.
   MAE is preferred when the focus is on the average magnitude of errors, and when we want a metric that's robust to outliers. Compared to MSE, MAE is less sensitive to outliers because it doesn't square the errors. This can be advantageous when dealing with data that has significant anomalies. MSE is preferred when it's crucial to penalize larger errors more and when optimization and mathematical properties are important.

1.2 Calculate the results of the measures that you have suggested for each simulation setting.

Done with the following sum of all MAE and MSE loss values for all students in each simulation setting. γ had a fixed value of 0.5.

Simulation 1: (57.20598619613622, 72.76614266591581)

Simulation 2: (56.575167193108335, 72.13532366288791)

Simulation 3: (60.24803256884263, 75.80818903862223)

Simulation 4: (30.767749504487856, 30.767749504487856)

Simulation 5: (37.727408126365106, 37.727408126365106)

1.3 Would changing the γ parameter change your measure of fit results and your decision of the best setting? (Comparing to $\gamma = 0.5$)

Yes, it does.
For $\gamma = 0.25$:

Simulation 1: (59.23215683859875, 74.79231330837833)

Simulation 2: (59.23215683859875, 74.79231330837833)

Simulation 3: (60.21029671978603, 75.77045318956563)

Simulation 4: (43.67200036881917, 43.67200036881917)

Simulation 5: (43.67200036881917, 43.67200036881917)

Got worse for all simulations.
For $\gamma = 0.75$:

Simulation 1: (60.24803256884263, 75.80818903862223)

Simulation 2: (60.04590922150732, 75.60606569128694)

Simulation 3: (60.24803256884263, 75.80818903862223)

Simulation 4: (28.963714875546927, 28.963714875546927)

Simulation 5: (17.259229732003597, 17.259229732003597)

Did much better for One-Skill Learning simulations.

1.4 Which simulation setting do you think is the best fit? Why?

I think One-Skill PFA learning simulations tend to be best fitting because they generally provide less MAE and MSE error summations than No-Skill PFA learning simulations do. The best modeling between the two mentioned, would be Low Learning One-Skill PFA due to robustness to variations of γ. Intuitively, considering skills in modelings provides more accurate predictions and forecastings of people's perfomance.

2 Plots

2.1 Box Plots

Box plots provide a visual summary of data distribution and make it easy to compare multiple datasets. By looking at a single box plot, we can see the median value, which is the central point of the data. The box shows the interquartile range (IQR), representing the middle 50% of the data, and the whiskers extend to indicate the range of the data. Any points outside the whiskers are considered outliers, which can highlight unusual values. When comparing multiple box plots, assessing the differences in medians to see which dataset has higher or lower central values is possible. We can also compare the spreads of the data by looking at the lengths of the boxes and whiskers. If the boxes and whiskers overlap, it suggests that the datasets have similar distributions. By examining the position and number of outliers, we can identify variations and inconsistencies between the datasets. Additionally, the relative position of the median within the box and the lengths of the whiskers can indicate whether the data is symmetric or skewed. Overall, box plots provide a comprehensive way to visualize and compare multiple datasets efficiently. In this case, comparing to the gold standard box plot, we observe how good is the distribution of calculated score is, and how many outliers are generated while calculating. These outliers can be considered as the mistakes and errors of calculations.
Although One-Skill PFA learning simulations provided least errors, but they also provided outliers. This is our due as the future observation and research. There may be some mistakes while implementing. The distributions of No-Skill PFA simulations are similar to the gold standard one and it makes sense more.

2.2 Bar Plots

A bar plot allows we to visualize categorical data by displaying rectangular bars, where the length of each bar is proportional to the value it represents. It's especially useful for comparing the frequencies, counts, or numerical values of different categories, making it easy to see which categories have higher or lower values. When comparing multiple bar plots, we can quickly assess differences and similarities across various categories. For example, comparing bar plots of sales data for different months can show we trends, highlight which products performed better in certain months, and reveal any significant changes over time. Comparing bar plots side by side can help we understand the relative importance or performance of categories in different contexts. A bar plot with 95% confidence intervals gives a visual representation of the central value for each category along with the range within which the true value is likely to fall 95% of the time. The bars show the average or total value, while the error bars (confidence intervals) show the uncertainty or variability in these estimates. When we compare multiple bar plots with 95% confidence intervals, we can see differences in central values across categories and understand the degree of uncertainty in these estimates. If the confidence intervals of two categories do not overlap, it suggests a significant difference between those categories, whereas overlapping intervals suggest that the difference may not be statistically significant. This helps in identifying trends, patterns, and statistical differences across categories, offering a comprehensive view of both central values and their reliability.

2.3 Histogram

A histogram plot gives we a visual representation of how data points are distributed across various intervals or bins. Including 95% confidence intervals in a histogram, it adds information about the reliability of the frequency counts in each bin. Essentially, it shows the range within which the true frequency counts are likely to fall 95% of the time. This helps in understanding the uncertainty or variability in the data distribution.

When comparing multiple histogram plots with 95% confidence intervals side by side, we can gain insights into differences and similarities in the distributions of different datasets. We can see how the data is spread across the bins for each dataset and compare the shapes of the distributions. The confidence intervals help we assess whether the observed differences in frequencies between datasets are statistically significant or if they might be due to random variation. This combined information allows we to better understand the consistency, variability, and potential trends or patterns within and across different datasets.

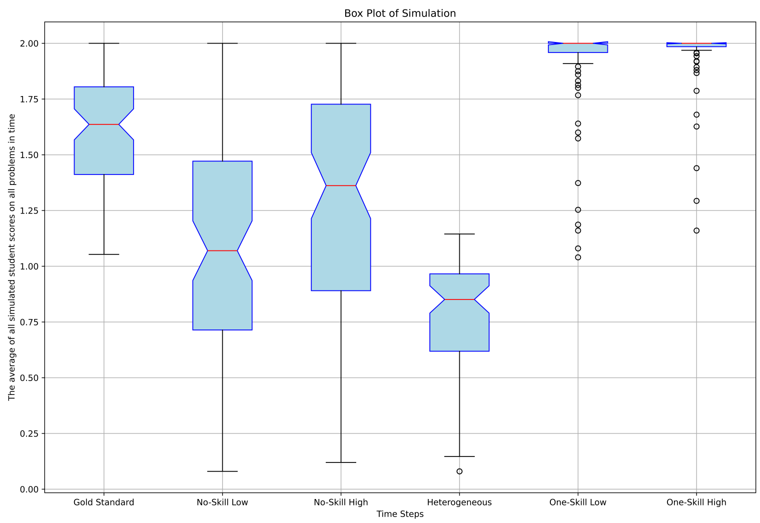

Figure 1: Box Plots

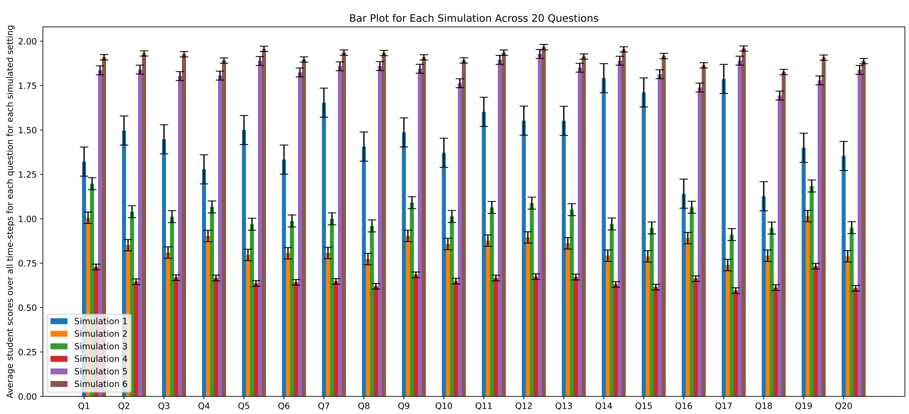

Figure 2: Bar Plots