Computer aided simulations and performance evaluation

Lab 3 - Peer Grading System Simulator

Gabriele Cuni 277957

3.1 Assumptions

The evaluation interval is an internal state and it is between zero and one.

$X_s \sim Uniform(0, 1)$ $\forall s \in S$	The quality of the student s is called X_s and it is uniformly distributed between zero and one..
$Q_{s,h} \sim Truncnorm(X_s, \sigma_Q, 0, 1)$ $\forall s \in S; \forall h \in H$	The quality of homework h of the student s is $Q_{s,h}$ that is a random variable. $Q_{s,h}$ has a truncated normal distribution between zero and one, mean equal to X_s and standard deviation equal to σ_Q
$E^{(k)}_{s,h} \sim Truncnorm(Q_{s,h}, \sigma_E, 0, 1)$ $\forall s \in S; K < S$ $\forall h \in H$	The evaluation k for homework h of the students is $E^{(k)}_{s,h}$. It is a random variable truncated normally distributed between zero and one, mean equal to $Q_{s,h}$ and standard deviation equal to σ_E
$\widehat{Q}_{s,h} = \frac{1}{K} \cdot \sum_{k \in K} E^{(k)}_{s,h}$ $\forall s \in S; K < S$ $\forall h \in H$	$\widehat{Q}_{s,h}$ is the estimator of the value of the homework h of the student s. Given that any student doesn't evaluate itself.

3.2 Input parameters

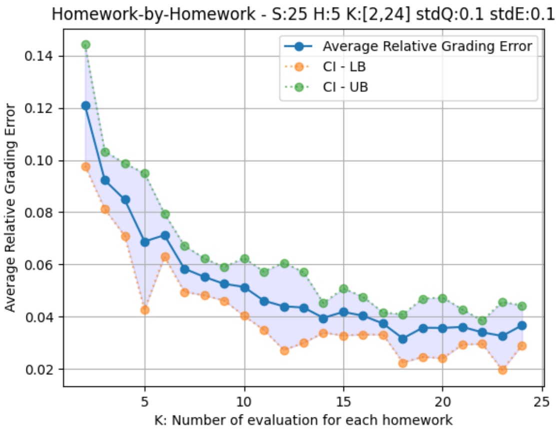
- S: number of students
- K: evaluators for each homework
- H: number of homework each student will make every simulation runs
- σ_Q: the quality variance that affects any homework submitted by any student.
- σ_E: the evaluation variance that affects any evaluation given by any student.

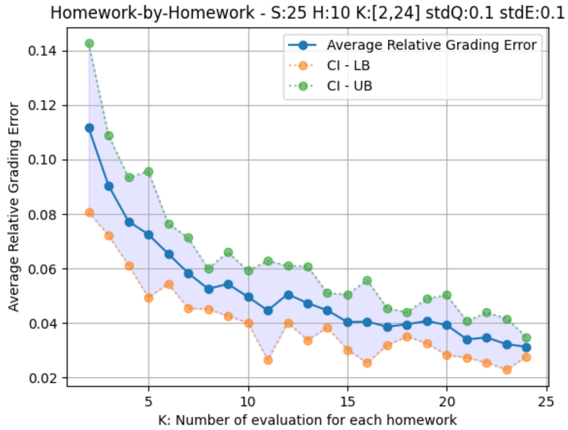
3.3 Output parameter

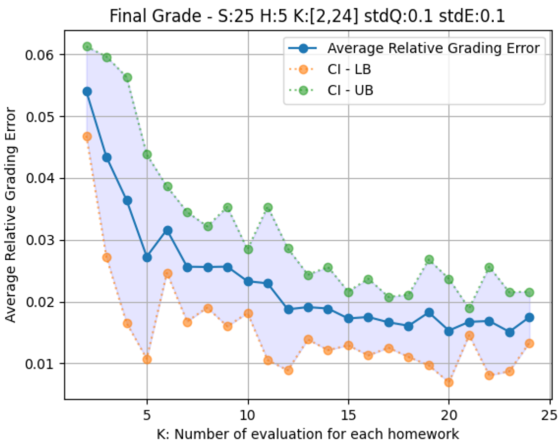
(homework-by-homework) average relative grading error $\epsilon_1 = \frac{1}{H \cdot S} \sum_{s \in S} \sum_{h \in H} \left| \frac{\widehat{Q}_{s,h} - Q_{s,h}}{Q_{s,h}} \right|$

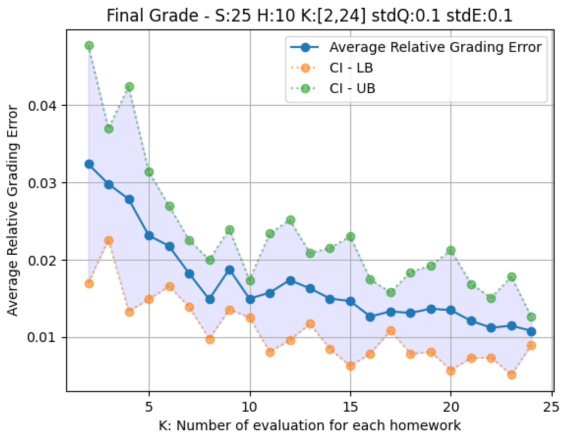
(final grade) average relative grading error $\epsilon_2 = \frac{1}{S} \sum_{s \in S} \frac{\left| \sum_{h \in H} (\widehat{Q}_{s,h} - Q_{s,h}) \right|}{\sum_{h \in H} Q_{s,h}}$

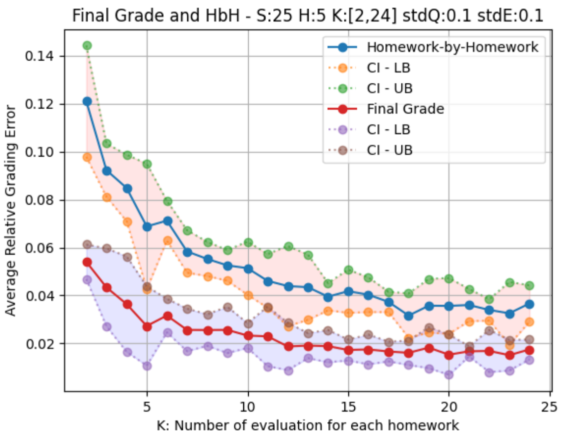
3.4 Charts

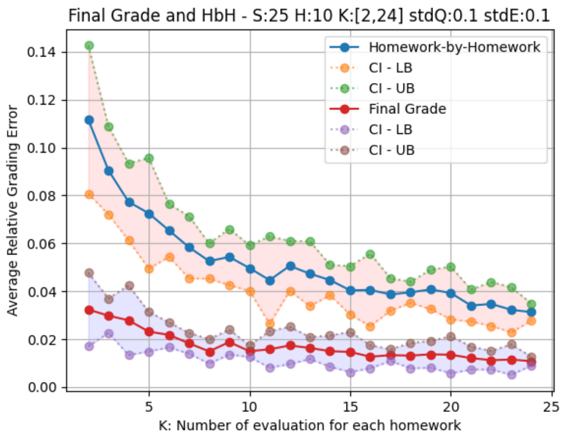

Homework-by-Homework - S:25 H:5 K:[2,24] stdQ:0.1 stdE:0.1
Average Relative Grading Error
CI - LB
CI - UB
Average Relative Grading Error
K: Number of evaluation for each homework

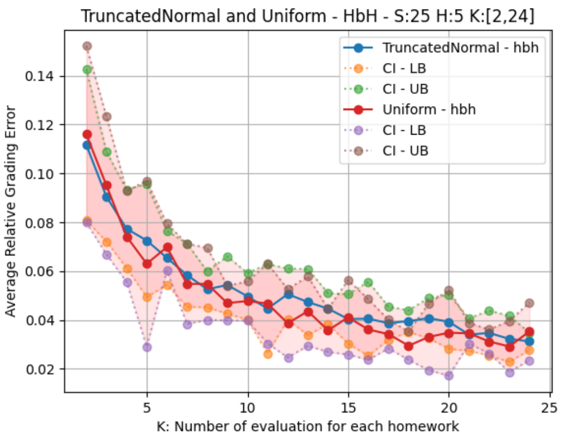

Homework-by-Homework - S:25 H:10 K:[2,24] stdQ:0.1 stdE:0.1
Average Relative Grading Error
CI - LB
CI - UB
Average Relative Grading Error
K: Number of evaluation for each homework

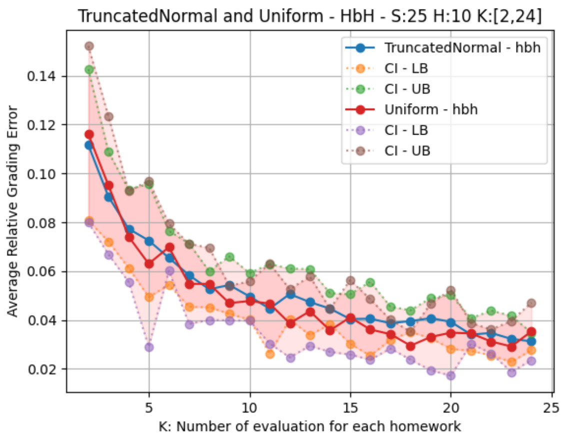

Final Grade - S:25 H:5 K:[2,24] stdQ:0.1 stdE:0.1
Average Relative Grading Error
CI - LB
CI - UB
Average Relative Grading Error
K: Number of evaluation for each homework


Final Grade - S:25 H:10 K:[2,24] stdQ:0.1 stdE:0.1
Average Relative Grading Error
CI - LB
CI - UB
Average Relative Grading Error
K: Number of evaluation for each homework


Final Grade and HbH - S:25 H:5 K:[2,24] stdQ:0.1 stdE:0.1
Homework-by-Homework
CI - LB
CI - UB
Final Grade
CI - LB
CI - UB
Average Relative Grading Error
K: Number of evaluation for each homework

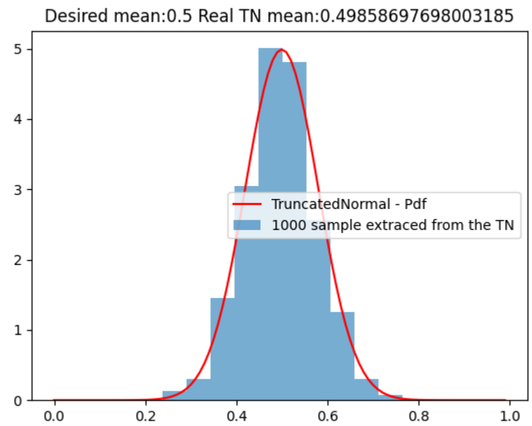

Final Grade and HbH - S:25 H:10 K:[2,24] stdQ:0.1 stdE:0.1
Homework-by-Homework
CI - LB
CI - UB
Final Grade
CI - LB
CI - UB
Average Relative Grading Error
K: Number of evaluation for each homework

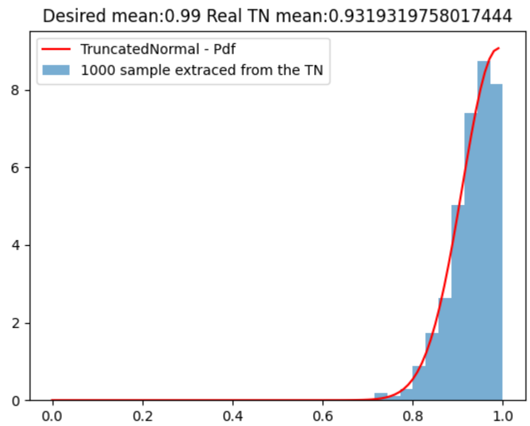

TruncatedNormal and Uniform - HbH - S:25 H:5 K:[2,24]
TruncatedNormal - hbh
CI - LB
CI - UB
Uniform - hbh
CI - LB
CI - UB
Average Relative Grading Error
K: Number of evaluation for each homework


TruncatedNormal and Uniform - HbH - S:25 H:10 K:[2,24]
TruncatedNormal - hbh
CI - LB
CI - UB
Uniform - hbh
CI - LB
CI - UB
Average Relative Grading Error
K: Number of evaluation for each homework

3.5 Observation about the algorithm

The key point to get the convergence of the average relative grading error is to recognize that extreme good or bad student quality led to an unexpected mean of the truncated normal distribution. For example, given a student with a very good quality as 0.99 and a standard deviation of 0.08 the truncated normal of the scipy library fails, because in order to guarantee the desired standard deviation the mean of the distribution shifts from the desired value to a lower value.





Given the above reasoning the mean is checked before use it to generate the distribution, if the value is too close at the extremes of the evaluation interval [0,1] the standard deviation σ_Q and σ_E are shrinked in order to guarantee the right expected mean.

3.6 Conclusions

The charts show that the average relative grading error converges to zero given any set of input parameters that I tested. This cannot be said of the confidence interval relative error that is not shown in any charts because it neither decreases nor increases by an appreciable quantity. Probably it is so because the number of runs each experiment is constant and equal to five.
The first fourth charts show that both ϵ_1 and ϵ_2 decrease by increasing the number of homeworks each student must make.
The fifth and sixth charts show that increasing the number of homeworks the distance between ϵ_1 and ϵ_2 increases also. In addition let H by 10, given that the confidence intervals do not intersect I can say that ϵ_1 is greater than ϵ_2 for the entire duration of the simulation.

Finally as optional modification I changed the probability distribution of $Q_{s,h}$ and $E^{(k)}_{s,h}$ from the truncated normal to the uniform distribution.
Given that also the uniform distribution requires the reasoning done in the previous section in order to properly work, the last two charts show that nothing can be said about what distribution is better to model the problem, given that the confidence intervals are overlapped.