Computer aided simulations and performance evaluation

Lab 5 - Random Dropping Implementation

Gabriele Cuni 277957

5.1 Assumptions

- The default scenario provides that the number of bins is equal to the number of balls and it is called n.
- The value d must be much smaller than n.
- X_i is the occupancy of the bin i.
- Given a confidence level, the number of runs must be chosen in order to have no overlapping confidence intervals, otherwise it is impossible to draw conclusions.
- The bin index is discrete uniform distributed between zero and $n - 1$.
- $theo = \frac{\log n}{\log\log n} \cdot (1 + O(1))$ is used when $d = 1$and the constant is assumed equal to one.
- $theo = \frac{\log\log n}{\log d} + O(1)$ is used when $d > 1$and the constant is assumed equal to zero.
- $theo_{UB} = 3 \cdot \frac{\log n}{\log\log n}$ is used when $d = 1$.
- The optional scenario provides that $noBalls = \alpha \cdot n$ where $n$is the number of bins and $d = 1$.
- The optional scenario provides that $theo = \alpha \cdot \frac{\log n}{\log\log n} \cdot (1 + O(1))$ when $d = 1$and the constant is assumed equal to one.
- The optional scenario provides that $theo_{UB} = 3 \cdot \alpha \cdot \frac{\log n}{\log\log n}$ is used when $d = 1$.

5.2 Input parameters

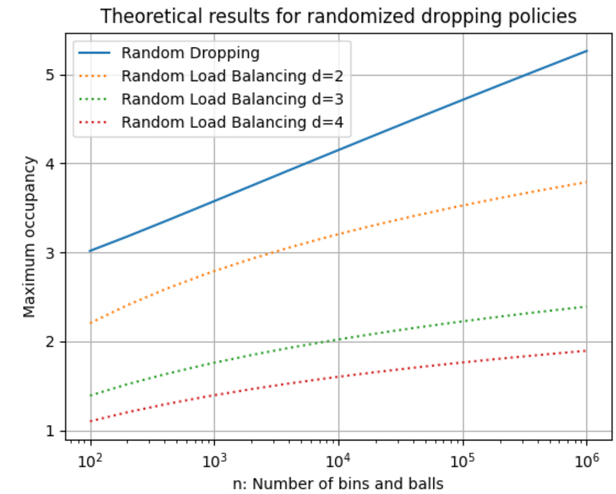
- **noBins**: Number of bins.
- **noBalls**: Number of balls.
- **d**: Number of bins that are selected in the Random Load Balancing Policy.
- **alpha**: Floating point number used to change the number of balls given the number of bins.
- **seed**: The number used to set the seed of the python random generator.
- **CI_CL**: Confidence level of the confidence interval.
- **noRuns**: Number of runs for each experiment. Any time I change the values of one input parameter I do a new experiment.

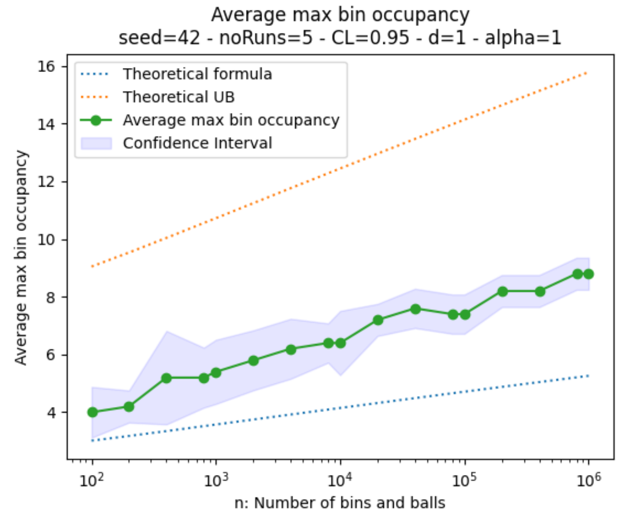
5.3 Output parameters

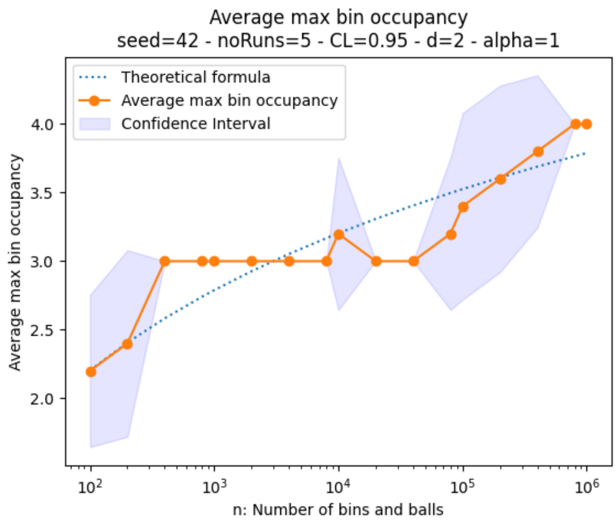
- **x_hat:** It is the average maximum occupancy and it is defined by $E(max(X_i))$
- **CI_LB**: Confidence interval lower bound.
- **CI_UB**: Confidence interval upper bound.
- **theo**: It is the value given by the theoretical maximum occupancy and it depends on the load balancing policy.

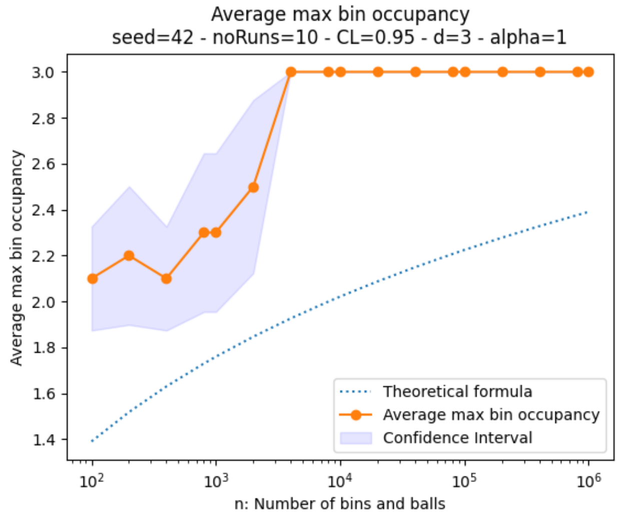
- **theo_UB**: It is the theoretical upper bound of the maximum occupancy given that the load balancing policy with d=1 is used. Otherwise It is meaningless.
- **CI_rel_err**: confidence interval relative error.

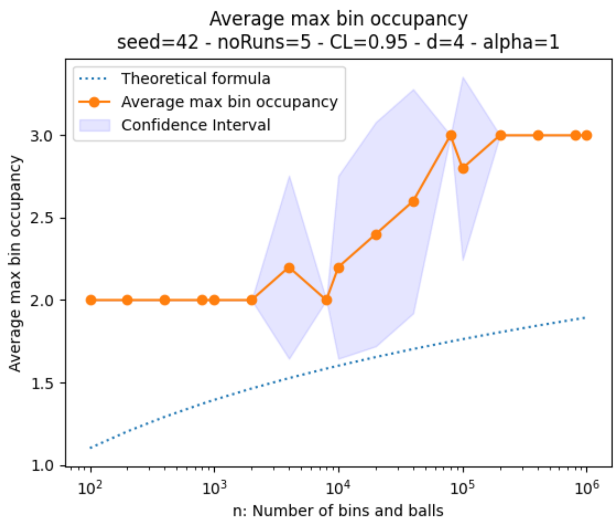
5.4 Charts

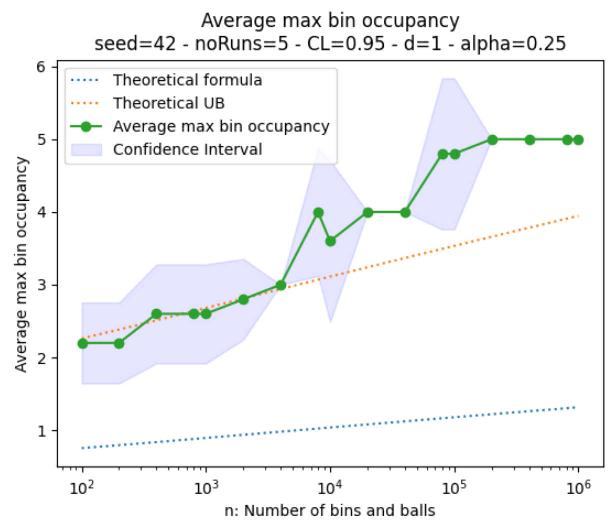


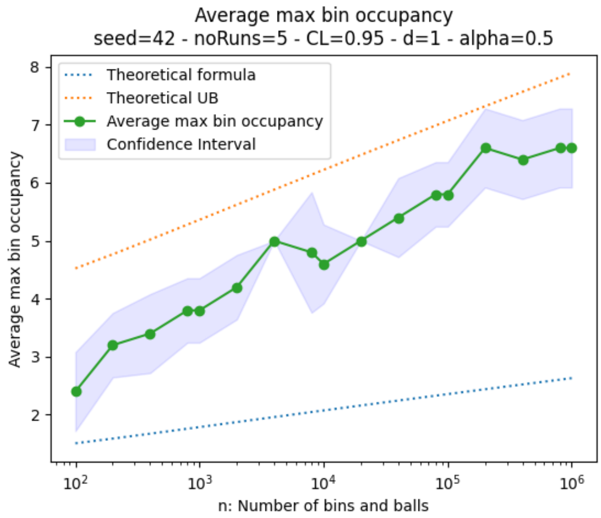


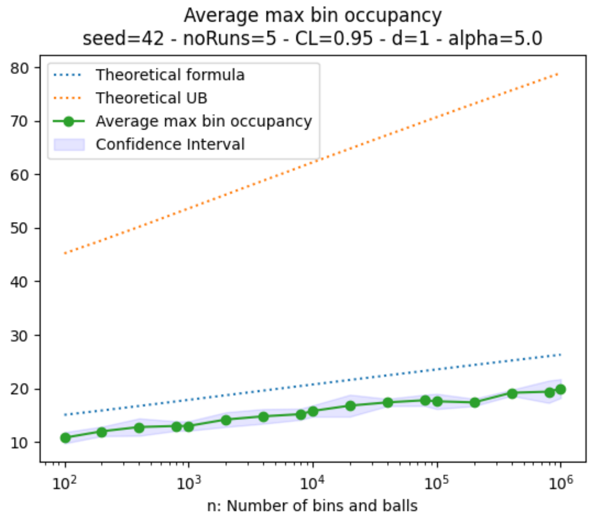












5.5 Conclusions

The second chart shows that 5 runs are enough to claim that the theoretical upper bound formula is right, actually it shows also the formula $theo = \frac{\log n}{\log\log n}$ is a lower bound given the simulation range.

The third chart shows how the simulated curve seems to follow the theoretical one given that for some points the confidence interval overlaps the theoretical curve.

The fourth charts shows that the formula $theo = \frac{\log\log n}{\log d}$ is a lower bound of the simulated one, actually 10 runs were needed to claim so, because the curves were overlapping with 5 runs.

Finally the fifth charts shows that if $d = 4$, five runs are enough to claim theo as a lower bound.

5.6 Optional scenario

The last three charts show the simulation results when the number of balls and bins are different. The theoretical formula and theoretical upper bound are modified as mentioned in the assumption section. It is shown, when $\alpha = 0.25$so that $noBalls < noBins$, the theoretical formula of the upper bound is not right. Nevertheless, when $\alpha = 5$so that $noBalls > noBins$, the theoretical formula is actually an upper bound instead of a lower bound.

Given these reasoning the proposed theoretical modification it is not able to satisfied the scenario in which $noBalls \neq noBins$.