A Simple Heuristic for Reducing the Number of Scenarios in Two-Stage Stochastic Programming

Abstract

- Two stage stochastic problems are often converted to deterministic multi scenario optimization problems, by discretizing the uncertain parameters in a finite set of scenarios.Since the total number of scenarios is the product of the number of discrete values each variable can take, problems can easily yield large scenario sets.
- The heuristic is meant to reduce the number of scenarios that you need to consider in order to estimate different parameters of a dataset with reasonable accuracy.
- My project focused on just the condensation of scenarios- how much the reduction, and the size of the problems for which it can be done, while ignoring the accuracy of the estimates for parameters.

The Algorithm

The objective of the algorithm is to reduce the number of scenarios that need to be considered, while still getting good estimates of different parameters of the dataset.

The basic idea was that the probability of a specific variable taking a specific value is the sum of probabilities of all the scenarios in which the concerned variable takes the corresponding value. The algorithm sets up a large linear program that redistributes the probabilities of scenarios (setting some to zero, hence “reducing” the total number) such that the probabilities of every value are still consistent with the initial conditions.

Specifics of the Program

1. Number of variables = 2 x Total number of Scenarios
2. Number of constraints = 1 + Number of Values + Total Number of Scenarios

   Note that: Total number of scenarios grows exponentially with increase in number of variables

All of the coding was done in Python using Numpy, Multiprocessing and the Python API of CPLEX

To test it, sample datasets were generated using Numpy, with uniform and Gaussian distributions, and a timeout of four hours was set.

Results

For Uniform Distributions-

- Worked almost always for upto 6 variables, and sparingly with upto 8, each with 2-5 values
- Large reduction in scenarios, starting from 50 -> 6, at best 3200 -> 8 (7 var)

For Gaussian Distributions-

- Worked almost always only for upto 3 variables, sparingly for 4 and 5, that too for a small number of total scenarios, and timed out/memory for 6 variables and beyond
- Large reductions, from 48 -> 9, and upto 96 -> 8

Conclusion

- The proposed heuristic can reduce the number of scenarios to less than 1% of the original number.
- The proposed heuristic, because of the large size of the linear program, results in a very large number of CPLEX constraints and variables for comparatively simple problems.
- As a result, the program requires unreasonable amounts of time and memory for most problems with anything above ten discretely distributed variables, meaning that it cannot be used for even slightly large datasets.