POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Sample Average Approximation for the Unit Commitment Problem with Stochastic Demand

Master Thesis in
Master in Renewables and Environmental Sustainability - Energy Engineering

Author: **Tomás Ramos Lara**

Student ID: 935900
Advisor: Prof. Ola Jabali
Co-advisors: Giovanni Pantuso
Academic Year: 2023-05

Abstract

In this thesis we study the Unit Commitment Problem with Stochastic Demand. We solve this problem using the Sample Average Approximation method with diverse instances, changing the deviation values and number of generators. We tested how this method performs when the amount of scenarios to be evaluated are too large to be handled by CPLEX. These tests were implemented in python with Gurobi solver. The results showed a considerable decrease in the time needed to find an acceptable solution to the problem. Moreover, the results demonstrated high quality of results.

Contents

1 | Introduction

1.1. Energy Systems and the UCP problem

The electricity power industry across the globe is experiencing a radical change in its business and operational models, undergoing restructuring and deregulation. Electricity providers are responsible for the generation, transmission and distribution of electricity, and must guarantee a reliable service of high quality. Since electricity cannot be easily stored, it's delivery is practically instantaneous; generation and demand are permanently in balance to keep the stability and integrity of the system. Therefore, power generating units are scheduled in advance in order to satisfy the upcoming forecasted demand, allowing preventive measures and action planning to cope with setbacks. Forecasting demand and scheduling production is of extreme importance to guarantee energy supply at all times. Thus, electricity providers face the challenging problem of deciding, from a set of electrical generators (i.e. generating units), **which power units** to operate (the unit status), in **which periods** and at **what level of production**, in order to satisfy the **demand of electrical energy**. This optimization problem is traditionally known in operation research as the **Unit Commitment Problem** (hereinafter **UCP**). The two main objectives of this problem are either to minimize cost or to maximize revenue, while respecting the constrains of the system.

The UCP is a critical task for the the operation of power systems, and its efficient solution offers many advantages to market players and final customers. Ideally, we seek to find optimal solution for this problem. However this is a challenging task considering the magnitude of the problem, its complicating constrains and possible computational limitations. For this reason, there are numerous studies in the literature where different approaches are proposed to find an optimal solution to this problem. This constitutes a fundamental objective for the progress of operational research in this field

The first UCP models were deterministic (e.g. Ahmad, Aijaz et al. [11]), problems were small and the demand to be satisfied was predictable, the systems were simpler than today's. As we will see, the applications of the UCP nowadays calls for new formulations that take into account for further complexities and the uncertainty of available data.

1.2. Uncertainty in the Power Systems

In past decades, coal-fired power plants were generally the thermal basis of power systems, while combined cycle gas turbines were relegated for high-demand periods and fast-ramping gas turbines were used to cover demand peaks. This operation was steady over time and did not demand further modeling developments to achieve an efficient operational management. Nowadays, the trend is changing: updated greenhouse policies, implementation of emission-allowance trading right markets, and many renewable-oriented political decisions have brought a paradigm shift. **Renewable Energy Sources** (**RES**) increasing penetration has brought new challenges to formulation of appropriate UCPs. For example, the inclusion of wind and photovoltaic energy sources in energy markets lead to uncertainty in production yield. Given that this depends on the weather, a variable extremely challenging to predict precisely, the resulting UCPs are extremly challenging.

In general, the operation of real power systems is implicitly subject to uncertainty. The RES production forecasts are highly dependant on weather and environmental variables, while demand is extremly sensitive to inaccuracy or sudden changes due to unexpected events. The continuous growth of electricity markets has made forecasting an increasingly complex and important challenge to be taken into account. For these reasons, the consideration of uncertainty in UCPs enhances the reliability of the resulting schedule. In this thesis we investigate a UPC under uncertainty.

There are several alternatives to transform a deterministic problem into a stochastic one (e.g. Louveazu et al. [3]). The most popular techniques applied to the unit commitment problem under uncertainty are Stochastic Optimization, Robust Optimization and Monte Carlo Simulation, which will be discussed in Section 2. The term stochastic refers to the property of a variable of being able to be represented by a random probability distribution. This distribution might be analysed statistically but can not be predicted precisely. Therefore, when using stochastic optimization, we are actually taking decision considering different realizations of future events, called scenarios, which have an associated realization probability.

The more uncertain a variable is, a larger number of scenarios needs to be considered in the model. In such cases, finding an exact solution becomes a challenging. Therefore, techniques have been developed to approximate the solution under less computational effort.

Uncertainty is found mainly in several parameters of the UCP. For example, the energy price (when the problem considers buying and selling electricity from the market), or in technical aspects, like maintenance operations. However, the main source of uncertainty in UCPs relates to the demand, as consumption is influenced by external factors which might cause it to rapidly increase or decrease. Demand's uncertainty affects schedule and raises new challenges in the context of the UCP. Therefore, various techniques and methods have been studied and employed to control the consequences of uncertainties associated with demand uncertainty. In this thesis, we propose a sample average approximation method for the **UCP under Stochastic Demand** (hereinafter **UCPSD**).

In section 2 we review the relevant the literature. In section 3 we describe our problem, and present our solution method in section 4. We describe our computational study in Section 5 and show its results in Section 6. Lastly, we present our conclusions in Section 7.

2 | Literature Review

The unit commitment problem is frequently addressed as an optimization problem with the objective of minimizing costs (e.g. Correa, Augusto [22]).The problem is sometimes modeled as a profit-maximization problem (e.g. Abdi, Hamdi [1]). Nowadays profit-based UCPs are gaining importance given the privatization and restructuring process that has been taking place in the energy power industry.

The solution methods used to solve the UCP can be categorised as classical/conventional approaches, non-classical approaches and hybrid techniques (Mallipeddi et al. [13]). Classical algortihms are the deterministic ones, where the most know approaches are Dynamic Programming (e.g. Padhy [18]), Branch and Bound (e.g. Pales et al. [19]), Lagrange Relaxation (e.g. Ongsakul et al. [16] or Shiina [25]) and Mixed Integer Lienar Programming (e.g. Xie et al. [30]).

In the last decade, there has been a significant growth in the application of non-classical approaches, mainly there has been a significant development of stochastic models, as researchers have observed that stochastic models perform better than deterministic models under uncertainty (e.g. Takriti et al.(1996) [27]). As previously mentioned, the changes that the global electric power sector has been facing have increased the uncertainty associated with various input parameters of UCPs. Different studies and reviews were published considering uncertainty management, attempting to control the consequences of uncertainties associated with parameters.

There are several approaches to cope with uncertainty. Namely, Two-Stage Stochastic Programming (e.g. Geng et al. [6], Huand et al. [8] and Wang et al. [29]) and Multi-Stage Stochastic Programming (e.g. Shiina et al. [26] and Zou et al. [33]) the most common ones. These formulations are based on modeling a decision as a random experiment appropriately described by a probability space. The random parameters of the problem are described by a random variable whose value is populated by the outcome of the random experiment. When the random variable is discrete, it counts a finite number of realizations, which are called *scenarios* in the stochastic programming jargon. Other methods like Risk Consideration Stochastic Programming (e.g. Xiong et al. [31]) , Chance

Constrained Stochastic Programming (e.g. Wu et al. [32]) or other hybrids approaches (e.g. Sayed et al. [23]) are less common and solve particular formulations of the UCP. For further study on the mentioned topics please refer to Montero et al. (2022) [14] and Mallipeddi et al. [13], or to Martin Haberg's work (2019) [9] for specific analysis on stochastic models.

Exact mathematical programming methods are less common nowadays, as they are restricted to cases with reduced uncertainty, that's to say, small numbers of scenarios. In this thesis we propose a sampling based approach to handle a large number of scenarios. Over the years the sampling methods have been used in stochastic problems in various ways, from discrete optimization by Kleywegt [10] to programs with integer recourse by Ahmed et al. [2] and even to solve routing problems (e.g. Verweij et al. [28]). We will applied the **Sample Average Approximation Method** (**SAA**) in the UCPSD with uncertain demand.

The SAA method uses random discrete samples draw from the true distribution of the uncertain parameters to generate scenarios. The UCPSD problem is then solved for this sample instead of the original scenario set. Then it replicates the process over several iterations to estimate the solution. The quality of these estimates is assessed by an analysis on the optimality gap and a confidence intervals. There are several methods to generate a limited number of scenarios from either a specified continuous distribution or a large data set that describes the uncertain parameters. These include random sampling (e.g. Glasserman [7]), moment matching (e.g. Ponomareva et al. [20] and scenario reduction by distance measures (e.g. Dupacová et al. [5]). In the UCP, the first one and the last one are mainly used, while moment matching is more common in power generation expansion problems.

In this thesis, we present a computational study of the application of the SAA to solve a cost-minimization UCPSD. In implementing the SAA, we followed de Mello et al. algorithm [4], Pernille et al. [24] formulations and Verweij [28] methodology. The main objective of this thesis is to examine the efficiency of SAA for the UCPSD.

3 | The Unit Commitment Problem with Stochastic Demand

3.1. Problem Description

The problem's input is composed of some known data, e.g., technical parameters, and the distribution of the demand. Considering a finite discrete planning horizon denoted as $\mathcal{T}$, the problem is formulated in two decision stages. In the first stage, we decide which generators are going be turned on or committed to production across the entire planning horizon. In the second stage, the actual demand values are revealed at each time period, and we decide the production level of each unit in order to satisfy it. In the second stage, production decisions are made while being bounded by the commitment decisions realized in the first stage.

We considered demand to be a random parameter described by a random variable. Specifically, we model it as a **discrete random variable**, counting for a finite number of realizations called scenarios. Each scenario is a vector of $|\mathcal{T}|$ elements representing a possible realization of demand at each time period $t \in \mathcal{T}$. We will be using $\mathcal{S}$ to denote the set of scenarios and s for each scenario.

A generic formulation of this two stage stochastic problem is as follows:

$$\begin{aligned} &\min_{x \in X} z = C^T x + Q(x) \\ &\text{s.t.} \quad Ax = b \\ &x \geq 0 \end{aligned} \tag{3.1}$$

where $Q(x) = \mathbb{E}_{\mathcal{S}}[Q(x,s)]$ is referred to as the *recourse function* and $\mathbb{E}_{\mathcal{S}}[Q(x,s)]$ is the expectation of the second stage recourse cost over all scenarios $s \in \mathcal{S}$. The C^T represents the costs associated with the first stage decision variable x, and the matrix $Ax = b$

describes the constrains to be satisfied in the first stage problem.

For a given scenario $s \in \mathcal{S}$, $Q(x, s)$ is defined as an optimization problem corresponding to the second stage decisions where x shows up as a right hand side parameter and the objective is to minimize the total recourse costs associated with the second stage decision variables y:

$$\begin{aligned} &Q(x,s) = \min_{y \in Y} c^T y \\ &\text{s.t.} \quad Wy = h - Tx \\ &y \geq 0 \end{aligned} \tag{3.2}$$

Given this, $Q(x)$ represents the optimal objective function value of the second stage problem given a certain x. Thus, x is an input of Q and therefore, it is considered as a parameter in the second stage problem.

The variable y denotes the second-stage production levels decisions, while c^T the costs associated with the y decision variables. The expression $Wy = h - Tx$ describes the constrains to be satisfied in the second stage problem.

This entails that demand uncertainty is hidden in the *recourse function* $Q(x)$, given that first stage decision are made without knowing the demand values.

We will model demand as a discrete random variable using the Monte Carlo sampling method. Given that $\mathcal{S}$ counts for a finite number of scenarios, we can account for different y for each scenario s, therefore, y_s will be telling us what to do in case scenario s materializes. Then the expectation function of $\mathbb{E}_{\mathcal{S}}[Q(x, s)]$ can be approximated by:

$$\mathbb{E}_{\mathcal{S}}[Q(x,s)] = \sum_{s \in \mathcal{S}} q_s c_s^T y_s \tag{3.3}$$

where q_s corresponds to the probability of occurrence of demand scenario s and c_s^T the costs of the second stage variables given scenario s. In our case, c_s^T is independant of the realizations of demand and therefore, constant for all scenarios.

First stage decisions are made taking into account technical constraints and physical limitations into account. For example, the requirement of certain generators of staying on for a certain period of time after start-up before being able to be turned off again (called Minimum Up Time). Another example is that units should be turned off for at least a certain period of time (Minimum Down Time). Other parameters taken into in the first stage are the cost of committing a unit to production, which is the cost of keeping it on (independent of the quantity produced) and the cost of starting up a unit.

Second stage decisions or, production decisions, have to respect the commitment decisions made in the first stage, as well as some technical constraints, like upper and lower limits to power production. Other limits are the ramp up and ramp down limits, which constrains the increment or reduction of the power output between subsequent periods. Another decision to be made in the second stage is the shedding amount: if the committed units are not enough to satisfy the load, part of the demand could be shedded at a given cost in order to match production to the load.

3.2. Mathematical Model

Let $\mathcal{T} = \{1, \ldots, T\}$ be the set of time periods, $\mathcal{T}' = \{2, \ldots, T\}$ the set of time periods without taking into account the first time period $t = 1$, $\mathcal{G} = \{1, 2, \ldots, G\}$ be the set of generators and $\mathcal{S} = \{1, 2, \ldots, S\}$ be the set of possible scenarios for the uncertain data, in our case, demand. These sets are summarized in Table 3.1.

The term $u_{g,t}$ is a binary variable representing the state of unit $g \in \mathcal{G}$ at period $t \in \mathcal{T}$, meaning, $u_{g,t} = 1$ when the generator g is on and $u_{g,t} = 0$ when it's off. This variable is used to address the commitment cost and, more importantly, as input of the second stage problem ones the first stage is solved.

$$u_{g,t} = \begin{cases} 1, & \text{unit } g \text{ is on at time } t \ a \\ 0, & \text{otherwise} \\ & \forall g \in \mathcal{G}, \forall t \in \mathcal{T} \end{cases} \tag{3.4}$$

While $c_{g,t}$ is a binary variable representing if unit $g \in \mathcal{G}$ had been turned on at period $t \in \mathcal{T}$. This variable is mainly used to address the start-up cost in the objective function.

$$c_{g,t} = \begin{cases} 1, & \text{unit } g \text{ is was turned on at time } t \ a \\ 0, & \text{otherwise} \\ & \forall g \in \mathcal{G}, \forall t \in \mathcal{T} \end{cases} \tag{3.5}$$

Let $p_{g,t,s}$ be a continuous variable representing the power production of unit $g \in \mathcal{G}$ at period $t \in \mathcal{T}$ in scenario $s \in \mathcal{S}$. Finally, $l_{t,s}$ represents the amount of demand satisfied by shedding the load in period $t \in \mathcal{T}$ under scenario $s \in \mathcal{S}$.

For a summary of the variables refer to Table 3.2

Each scenario $s \in \mathcal{S}$ has an associated realization probability q_s, which represents how likely is it to get scenario s, and $\sum_{s\in S} q_s = 1$. In our case, we are working with a Monte Carlo simulation, so $q_s = \frac{1}{|S|}$ is constant and equal for each scenario. Lastly, the value for demand under the scenario $s \in \mathcal{S}$ for each period $t \in \mathcal{T}$ will be represented by $d_{t,s}$.

The last parameters are related to the initial conditions of each generator g, representing the state of the unit at the end of the previous scheduled period, denoted with $t = 0$ Let $uInit_g$, $pInit_g$ and $tInit_g$ be the state of generator g, its power production and the time periods that it has been on at the beginning of the scheduling horizon $t = 0$, for $g \in \mathcal{G}$. Note that negative values of $tInit_g$ account for the time periods the unit has been off at the beginning of the scheduling horizon.

Parameters C_g^S, C_g^C, C_g^P represents the Start Up, Commitment and Production costs respectively for generator $g \in \mathcal{G}$. Commitment cost is the cost associated to keeping a generator producing for a period (like costs associated to maintenance). We also have the production upper and lower bound P_g^{max} and P_g^{min} respectively, and let R_g^{Up} and R_g^{Up} be the ramp-up and ramp-down limitations for generator g. These values represent the maximum variation the production can have between periods. Lastly let T_g^{Up} and T_g^{Down} be the minimum up time and downtime respectively. L_t represents the cost of load shedding.

Parameters can be found in Table 3.3.

3.2.1. Formulation

We formulate the problem as follows:

$$\min \sum_{g\in G}\sum_{t\in T}\left(C_g^S c_{g,t} + C_g^C u_{g,t} + \sum_{s\in S} q_s \cdot \left(L_t l_{t,s} + C_g^P p_{g,t,s}\right)\right) \tag{3.6}$$

subject to

$$c_{g,1} \geq (u_{g,1} - uInit_g) \qquad \forall g \in \mathcal{G} \tag{3.7}$$

$$c_{g,t} \geq (u_{g,t} - u_{g,t-1}) \qquad \forall g \in \mathcal{G},\ \forall t \in \mathcal{T}' \tag{3.8}$$

$$\sum_{g\in G} p_{g,t,s} + l_{t,s} \geq d_{t,s} \qquad \forall t \in \mathcal{T},\ \forall s \in \mathcal{S} \tag{3.9}$$

$$p_{g,t,s} \geq P_g^{Min} u_{g,t} \qquad \forall g \in \mathcal{G},\ \forall t \in \mathcal{T},\ \forall s \in \mathcal{S} \tag{3.10}$$

$$p_{g,t,s} \leq P_g^{Max} u_{g,t} \qquad \forall g \in \mathcal{G},\ \forall t \in \mathcal{T},\ \forall s \in \mathcal{S} \tag{3.11}$$

$$p_{g,1,s} - pInit_g \leq R_g^{Up} \qquad \forall g \in \mathcal{G},\ \forall s \in \mathcal{S} \tag{3.12}$$

$$p_{g,t,s} - p_{g,t-1,s} \leq R_g^{Up} \qquad \forall g \in \mathcal{G},\ \forall t \in \mathcal{T}',\ \forall s \in \mathcal{S} \tag{3.13}$$

$$pInit_g - p_{g,1,s} \leq R_g^{Down} \qquad \forall g \in \mathcal{G},\ \forall s \in \mathcal{S} \tag{3.14}$$

$$p_{g,t-1,s} - p_{g,t,s} \leq R_g^{Down} \qquad \forall g \in \mathcal{G},\ \forall t \in \mathcal{T}',\ \forall s \in \mathcal{S} \tag{3.15}$$

$$\sum_{\delta=0}^{T_g^{Up}-1} (u_{g,\delta}) \geq T_g^{Up} c_{g,1} \qquad \forall g \in \mathcal{G} \tag{3.16}$$

$$\sum_{\delta=0}^{T_g^{Up}-1} (u_{g,\delta+t}) \geq T_g^{Up}(u_{g,t} - u_{g,t-1}) \qquad \forall g \in \mathcal{G}, \forall t \in \{2,...,T - T_g^{Up} + 1\} \tag{3.17}$$

$$\sum_{\delta=0}^{T_g^{Down}-1} (1 - u_{g,\delta}) \geq T_g^{Down}(uInit_g - u_{g,1}) \ \forall g \in \mathcal{G} \tag{3.18}$$

$$\sum_{\delta=0}^{T_g^{Down}-1} (1 - u_{g,\delta+t}) \geq T_g^{Down}(u_{g,t-1} - u_{g,t}) \forall g \in \mathcal{G}, \forall t \in \{2,...,T - T_g^{Down} + 1\} \tag{3.19}$$

$$u_{g,t}, c_{g,t} \in \{0,1\} \qquad \forall g \in \mathcal{G}, \forall t \in \mathcal{T} \tag{3.20}$$

$$p_{g,ts}, l_{t,s} \geq 0 \qquad \forall g \in \mathcal{G}, \forall t \in \mathcal{T}, \forall s \in \mathcal{S} \tag{3.21}$$

Indices	Description	List
$\mathcal{G}$	Set of generators	$\{1, 2, \ldots, G\}$
$\mathcal{T}$	Set of time periods	$\{1, 2, \ldots, T\}$
$\mathcal{T}'$	Set of time periods without $t = 1$	$\{2, 3, \ldots, T\}$
$\mathcal{S}$	Set of scenarios	$\{1, 2, \ldots, S\}$

Table 3.1: Notation for the sets

Variable	Description	Type
$\boldsymbol{u}_{g,t}$	Generator g status at time t	binary
$\boldsymbol{c}_{g,t}$	if generator g was turned on at time t	binary
$\boldsymbol{p}_{g,t,s}$	power produced by generator g at time t in scenario s	continuous
$\boldsymbol{b}_{t,s}$	amount of shedded energy at time t in scenario s	continuous

Table 3.2: Notation for the variables

Parameters	Description
$\boldsymbol{C}_g^S$	Start up cost for generator g
$\boldsymbol{C}_g^C$	Commitment cost for generator g
$\boldsymbol{C}_g^P$	Production cost for generator g
$\boldsymbol{B}_t$	Price of electricity at period t
$\boldsymbol{P}_g^{Min}$	Minimum production Level for generator g
$\boldsymbol{P}_g^{Max}$	Minimum production Level for generator g
$\boldsymbol{R}_g^{Up}$	Maximum ramp up time for generator g
$\boldsymbol{R}_g^{Down}$	Maximum ramp down time for generator g
$\boldsymbol{T}_g^{Up}$	Minimum up time for generator g
$\boldsymbol{T}_g^{Down}$	Minimum down time for generator g
$\boldsymbol{d}_{t,s}$	Energy demand at period t in scenario s
$\boldsymbol{q}_s$	Realization probability of scenario s
$\boldsymbol{uInit}_g$	State of generator g at start of the schedule
$\boldsymbol{pInit}_g$	Production level of generator g at start of the schedule

Table 3.3: Notation for the parameters

The objective of formulation (3.6) is to minimize the total costs, i.e., the total start up, commitment and production costs of all generators g added to the cost of shedded energy, along the time schedule $\mathcal{T}$, taking into account the possible scenarios s. Notice that the contribution of the second stage costs are expressed as a linear combination of the costs associated to each scenario s, expressed as $(L_t l_{t,s} + C_g^P p_{g,t,s})$ with constant probability realization of that scenario q_s.

The first two constraints (3.7) and (3.8) are called *associating constraints* and are responsible of populating the binary variable $c_{g,t}$ with 1 value when the status of generator g ($u_{g,t}$), changes from 0 to 1 at period t. The difference between these constraints is that (3.7) takes into account the condition of generator g before the scheduling horizon at $t = 0$ in order to populate just $c_{g,1}$, while (3.8) encompass the rest of the time horizon populating $c_{g,t}$ for $t \in \mathcal{T}'$.

Constraints (3.9) are called the *power balance constraints* and ensures that the power generated by all generators at a time period t for a scenario s meets the forecasted demand $d_{t,s}$. In this model there are two particular situations that must be taken into account. First, there is the possibility of over production: given the first-stage decisions, it could happen that we have an overproduction of energy compared to demand that cannot be compensated by reducing production given the ramping constraints or the production limit of the committed units. In general, an extra variable is added to account for the excess energy sold to the grid. In our case, we will assume the excess energy to be gifted to the grid at price zero and bear the cost of producing that energy by adding a "$\geq$" sign on the demand constraints. Other situation to be taken into account is underproduction. For this we added the $l_{t,s}$ variable which is used to address the cost of shedding part of the load to accommodate for the under production. This variable counts the energy reduction in demand by load shedding. The objective of this thesis is to analyze the efficiency of a particular solution methodology, so this shedding of energy allows the model to always have a feasible solution.

Constraints (3.10) and (3.11) restrain the *range of power generation* so that it fits the minimum and maximum production levels, respectively P_g^{Min} and P_g^{Max} for every generator g at all time periods t and scenarios s.

Constraints (3.12) to (3.15) restrain the variation in power generation of generator g between subsequent time periods for all time periods t in all scenarios s. The first two (3.12) and (3.13) are called *ramp-up constraints* and limit the increase of power out put of generator g between subsequent time periods to R_g^{Up}. While (3.14) and (3.15) are called *ramp-down constraints* and limit the decrease of power out of generator g between subsequent time periods to R_g^{Down}.

Constraints (3.16) to (3.19) are referred to as *minimum time constraints*, and are associated with the minimum up and down time of the generators. The first two, (3.16) and (3.17), ensures that every generator g in every scenario s will be committed (on) continuously for a certain time period T_g^{Up} before its decommitment (shutdown). While the other two, (3.18) and (3.19), ensures that every generator g in every scenario s will be decommited (off) continuously for a certain time period before its commitment.

We decided to create constraints (3.16) only for the case where generator g was off before the $t = 1$, i.e., for $uInit_g = 0$. If this condition applies, we can have two possible situations for $t = 1$: either generator g becomes on and $u_{g,1} = 1$, or it stays off and $u_{g,1} = 0$. In the first sitation, constraints (3.16) become active as $c_{g,1} = 1$ and forces the next T_g^{Up} periods to keep generator g on. On the other hand, if $u_{g,1} = 0$, we assumed that $t = 1$ was the last period of the required minimum down time T_g^{Down}. Therefore, there is no constraint for generator g to stay off for the next time periods. In this situation, we will have that $c_{g,1} = 0$ and constraints (3.16) do not become active. The same logic applies for the case where $uInit_g = 1$: constraints (3.18) are then generated and only may become active if $u_{g,1} = 0$, whereas if $u_{g,1} = 1$, we assume that it is the last period of the required T_g^{Up} and there is no need for generator g to stay on the next time periods.

When considering $t = 1$ the last period of the minimum up or down time limitation, we are ignoring the time periods that generator g was on or off before the time schedule $tInit_g$, and only considering the status of generator g at the beginning of the time horizon $uInit$. This simplification could slightly change the objective values, but likely not affect the overall conclusion of this thesis. In the worst case scenario, generator g could be turned off before T_g^{Up} time periods have passed (or turned on before T_g^{Down} limitation applies) which would constrain a little the solution, but would not make a significant impact.

4 | Sample Average Approximation Method

The **Sample Average Approximation (SAA)** method works by repeatedly solving the two-stage model previously formulated with a limited number of scenarios, sampled from the set of the true scenario set $\mathcal{S}$. In this technique, the expected objective function value of the stochastic problem is approximated by a sample average estimates derived from random samples. Below we provide a step wise procedure for the SAA algorithm based on Pernille et al. [24]. In this paper, sampled scenarios are generated by the Monte Carlo sampling method.

A sample is constructed by $w^1, w^2, \ldots, w^N$ of N *sample scenarios*, randomly generated from the set $\mathcal{S}$. We call N the size of the sample and q_w the realization probability for each scenario w^i in the sample. Given that we are working with Monte Carlo simulation, we know every scenario has the same probability, i.e., $q_w = \frac{1}{N} = constant$ for $w \in \mathcal{W} = \{w^1, w^2, \ldots, w^N\}$.

The resulting sample average approximating problem is then solved for sample set $\mathcal{W}$ instead of the whole set $\mathcal{S}$. We do so by solving the resulting deterministic extensive formulation in order to obtain an optimal value z_N and optimal solution $\hat{x}$ and $\hat{y}$. These will be used to provide estimates of the actual optimal value of z^*. For clarification, in our case $\hat{x}$ represents the first-stage variables $c_{g,t}$ and $u_{g,t}$

$$\hat{x} = (c_{g,t}, u_{g,t})_{g\in G, t\in T}$$

The **Sample Average Approximation problem** corresponding to the original two-stage stochastic problem stated in Section 3.2.1 can now be formulated in its deterministic equivalent problem as follows:

$$\min z_N = \sum_{g\in G}\sum_{t\in T}\left(C_g^S c_{g,t} + C_g^C u_{g,t} + Q(u)\right) \tag{4.1}$$

subject to

$$c_{g,0} \geq (u_{g,0} - uInit_g) \qquad \forall g \in \mathcal{G} \tag{4.2}$$

$$c_{g,t} \geq (u_{g,t} - u_{g,t-1}) \qquad \forall g \in \mathcal{G},\ \forall t \in \mathcal{T}' \tag{4.3}$$

$$\sum_{\delta=0}^{T_g^{Up}-1} (u_{g,\delta}) \geq T_g^{Up}(u_{g,0} - uInit_g) \qquad \forall g \in \mathcal{G} \tag{4.4}$$

$$\sum_{\delta=0}^{T_g^{Up}-1} (u_{g,\delta+t}) \geq T_g^{Up}(u_{g,t} - u_{g,t-1}) \qquad \forall g \in \mathcal{G}, 0\forall t \in \{1, ..., T - T_g^{Up} + 1\} \tag{4.5}$$

$$\sum_{\delta=0}^{T_g^{Down}-1} (1 - u_{g,\delta}) \geq T_g^{Down}(uInit_g - u_{g,0}) \ \forall g \in \mathcal{G} \tag{4.6}$$

$$\sum_{\delta=0}^{T_g^{Down}-1} (1 - u_{g,\delta+t}) \geq T_g^{Down}(u_{g,t-1} - u_{g,t}) \forall g \in \mathcal{G}, \forall t \in \{1, ..., T - T_g^{Down} + 1\} \tag{4.7}$$

$$u_{g,t}, c_{g,t} \in \{0, 1\} \qquad \forall g \in \mathcal{G}, \forall t \in \mathcal{T} \tag{4.8}$$

$$\tag{4.9}$$

where $Q(u)$ represents the optimal objective value of the second stage problem given a certain $u_{g,t}$ over all scenarios $w \in \mathcal{W}$.

As mentioned before, this function $Q(u)$ is the expectation of the second stage problem

$$Q(u) = \mathbb{E}_{\mathcal{W}}[Q(u, w)] \tag{4.10}$$

Given that we are sampling with Monte Carlo technique, the expectation can approximated by:

$$\mathbb{E}_{\mathcal{W}}[Q(u, w)] = \frac{1}{N} \sum_{w \in \mathcal{W}} (L_t l_{t,w} + C_g^P p_{g,t,w})$$

Now we can formulate $Q(u, w)$ as an optimization problem corresponding with the second stage decisions:

$$Q(u,w) = \min \sum_{g\in G}\sum_{t\in T} \left(L_t l_{t,w} + C_g^P p_{g,t,w}\right) \tag{4.11}$$

subject to

$$\sum_{g\in G} p_{g,t,w} + l_{t,w} \geq d_{t,w} \qquad \forall t \in \mathcal{T} \tag{4.12}$$

$$p_{g,t,w} \geq P_g^{Min} u_{g,t} \qquad \forall g \in \mathcal{G},\ \forall t \in \mathcal{T} \tag{4.13}$$

$$p_{g,t,w} \leq P_g^{Max} u_{g,t} \qquad \forall g \in \mathcal{G},\ \forall t \in \mathcal{T} \tag{4.14}$$

$$p_{g,0,w} - pInit_g \leq R_g^{Up} \qquad \forall g \in \mathcal{G} \tag{4.15}$$

$$p_{g,t,w} - p_{g,t-1,w} \leq R_g^{Up} \qquad \forall g \in \mathcal{G},\ \forall t \in \mathcal{T}' \tag{4.16}$$

$$pInit_g - p_{g,0,w} \leq R_g^{Down} \qquad \forall g \in \mathcal{G} \tag{4.17}$$

$$p_{g,t-1,w} - p_{g,t,w} \leq R_g^{Down} \qquad \forall g \in \mathcal{G},\ \forall t \in \mathcal{T}' \tag{4.18}$$

$$p_{g,t,w}, l_{t,w} \geq 0 \qquad \forall g \in \mathcal{G},\ \forall t \in \mathcal{T} \tag{4.19}$$

Notice that in constraints (4.13) and (4.14) we find the first stage decisions $u_{g,t}$ which are now parameters (and not variables anymore).[1]

This procedure is then repeated by generating M samples and solving several associated optimization problems to obtain candidate solutions along with statistical estimates of their optimality gaps.

4.1. Methodology

The SAA method consists of solving the SAA problem (4.1) several times for M independent samples, each composed of N scenarios, in order to generate the associated objective values $z_N^1, z_N^1, \ldots, z_N^M$ and their corresponding candidate solutions $\hat{x}_N^1, \hat{x}_N^2, \ldots, \hat{x}_N^M$ and $\hat{y}_N^1, \hat{y}_N^2, \ldots, \hat{y}_N^M$. These values are now used to get valuable information on the actual objective function value z^*

4.1.1. Lower Bound Estimate

Once we have generated M independent samples, each of composed of N scenarios, and solved the UCPSD problem M times for each candidate sample, we will have M optimal solutions z_N. We denote an optimal solution for a sample $m \in \{1,\ldots,M\}$ by z_N^m. We calculated the average of the optimal objective function values of the M SAA problems,

[1] This is just a representation to clarify the formulation of the second stage problem. In practice we solved the extensive formulation described in Section 3.2.1 considering $\mathcal{W}$.

which we will denote $\overline{z}_N$:

$$\overline{z}_N = \frac{1}{M}\sum_{m=1}^{M} z_N^m \tag{4.20}$$

Then $\mathbb{E}[\overline{z}_N] \leq z^*$, as proved in Mak et al. (1999) [12] and in Norkin (1998) [15]. Therefore, $\overline{z}_N$ provides a statistical estimate for a **lower bound (LB)** of the optimal value of the original problem $LB_{M,N}$:

$$LB_{M,N} = \overline{z}_N \tag{4.21}$$

The variance of the lower bound $\hat{\sigma}^2_{LB_{M,N}}(M)$ is estimated by the variance estimator:

$$\hat{\sigma}^2_{LB_{M,N}} = \frac{1}{(M-1)}\sum_{m=1}^{M}(z_N^m - \overline{z}_N)^2 \tag{4.22}$$

For this calculated estimate, we would like to know how much we expect to get close to the same estimate if we run again the SAA with different samples. This is called the **confidence interval** of an estimate and it measures the degree of uncertainty of a variable in a sampling method. It is a range of values, bounded above and below the statistic mean, providing lower bound and upper bound to the estimate, with a confidence level representing the percentage of probability that this interval would contain the solution value when a random sample is drawn many times. In this chapter we will be using the formulas for the confidence interval provided by Kleywegt et al. (2002) [10]:

$$\left[LB_{M,N} - z_\alpha \frac{\hat{\sigma}^2_{LB_{M,N}}}{\sqrt{M}}; LB_{M,N} + z_\alpha \frac{\hat{\sigma}^2_{LB_{M,N}}}{\sqrt{M}}\right] \tag{4.23}$$

Where z_α represents the critical value of the normal distribution for a confidence level of $1-\alpha$.

4.1.2. Upper Bound Estimate

For any candidate solution $\hat{x}_N^m = (\hat{c}_{g,t}^m, \hat{u}_{g,t}^m)$, the objective value $\sum_{g\in G}\sum_{t\in T}(C_g^S \hat{c}_{g,t}^m + C_g^C \hat{u}_{g,t}^m + \mathbb{E}[Q(\hat{u})])$ is an upper bound for z^*, since $\hat{x}_N^m$ is a feasible point of the true problem. This upper bound value is estimated by fixing the first-stage solution and solving the formulation for sample $\mathcal{W}'$ of size N' scenarios:

$$\hat{z}_{N'}(\hat{c}_{g,t}, \hat{u}_{g,t}) = \min \sum_{g \in G} \sum_{t \in T} \left(C_g^S \hat{c}_{g,t}^m + C_g^C \hat{u}_{g,t}^m + \frac{1}{N'} \sum_{w \in \mathcal{W}'} Q(u, w) \right) \tag{4.24}$$

For any of the feasible solutions $\hat{x}_N^m$ and $\hat{y}_N^m$, the objective value that comes from fixing the first stage variables on (4.1), and solving the problem, provides an upper bound on z^*. We can follow any criteria, we choose the solution that provides the smallest $\tilde{z}_N^m$.

The N' (called **reference sample size**) represent the size of the new sample $\mathcal{W}'$. We choose N' randomly from $\mathcal{S}$ and $N' \gg N$, i.e., quite larger than N. Ideally, we wish this reference sample to be the true distribution, but typically this is not possible. Therefore we choose it as close as possible to $|\mathcal{S}|$. Given that $\mathcal{W}'$ is randomly generated, we have an unbiased estimator, and therefore we have that $\mathbb{E}[\hat{z}_{N'}] \geq z^*$, providing a statistical estimate for an **upper bound (UB)** of z^*.

$$UB_{N'}(\hat{x}_N^m) = \hat{z}_{N'}(\hat{c}_{g,t}, \hat{u}_{g,t}) \tag{4.25}$$

The variance of the upper bound $\hat{\sigma}^2_{UB_{N'}}$ is estimated by the variance estimator:

$$\hat{\sigma}^2_{UB_{N'}} = \frac{1}{(N'-1)} \sum_{w \in \mathcal{W}'} \Big(\big(P' + Q(\hat{u}, w)\big) - \hat{z}_{N'}(\hat{c}_{g,t}, \hat{u}_{g,t}) \Big)^2 \tag{4.26}$$

where $P' = \sum_{g \in G} \sum_{t \in T} (C_g^S \hat{c}_{g,t}^m + C_g^C \hat{u}_{g,t}^m)$ of the given candidate first-stage solution and $Q(\hat{u}, w)$ represents the optimal solution of the second-stage for scenario w for a given first stage optimal solution $\hat{u}$.

As mentioned before, for the upper bound we also need to calculate a confidence level, using the same formula:

$$\left[UB_{N'}(\hat{x}_N^m) - z_\alpha \frac{\hat{\sigma}^2_{UB_{N'}}}{\sqrt{N'}} ; UB_{N'}(\hat{x}_N^m) + z_\alpha \frac{\hat{\sigma}^2_{UB_{N'}}}{\sqrt{N'}} \right] \tag{4.27}$$

4.2. Optimality Gap Estimate

Once we have calculated our estimates and confidence intervals, the most important question we need to ask is *how close is z^* to these upper and lower bounds?* Which would be the equivalent of asking, how well our samples perform in comparison with the original scenarios in finding a candidate solution? To do this, we would like to compute **optimality**

gaps, defined as the distance between the estimate and the best known solution.

In our case, it would be formulated as UP - z^* and z^* - LB. Unfortunately, the very reason to develop the methodology described in this paper is that the computation of this solution z^* is extremly hard. For this reason, we use the proposed formulations by Seljom et al. [24] of the estimator of the optimality gap $GAP_{M,N,N'}$, its variance $\sigma^2_{GAP_{M,N,N'}}$ and the confidence interval of this gap based on the calculated optimality gaps of the estimated bounds, according to the following equations:

$$GAP_{M,N,N'} = UB_{N'} - LB_{M,N} \tag{4.28}$$

$$\sigma^2_{GAP_{M,N,N'}} = \frac{\hat{\sigma}^2_{UB_{N'}}}{\sqrt{N'}} + \frac{\hat{\sigma}^2_{LB_{M,N}}}{\sqrt{M}} \tag{4.29}$$

$$\left[GAP_{M,N,N'} - z_\alpha \sigma^2_{GAP_{M,N,N'}}; GAP_{M,N,N'} + z_\alpha \sigma^2_{GAP_{M,N,N'}} \right] \tag{4.30}$$

4.3. Example

We present a small example to help understanding the methodology. We used the data provided for the model (see Appendix A) applying the SAA method for a set composed of 20 scenarios $|\mathcal{S}| = 20$ solved for 10 independent samples $M = 10$ composed of $N = 3$ scenarios each. For the upper bound, we assume $N' = S = 20$, as S is relatively small.

We used an instance of $G = 12$ generators and a standard deviation σ of 15%. This σ is taken into account when sampling the demand to generate the scenarios. For each time period $t \in \mathcal{T}$ we define its standard deviation as a percentage of the average demand of that time period t, i.e., $\sigma_t = \sigma d_t$.

Finding the exact solution is relatively easy:

$$z^* = 525,754$$

For the lower bound, we generate $M = 10$ independant random samples and solve the UCPSD problem for each of them.

The expected value of the average is calculated by simply multiplying each objective value by its probability, which is $q_m = \frac{1}{10}$, and added to the product.

Sample	Objective Value
1	512,871
2	547,378
3	551,181
4	510,364
5	509,150
6	517,043
7	501,330
8	520,173
9	513,322
10	519,851

Table 4.1: Example: UCPSD Objective Values

$$\overline{z}_N = 520,266$$

Following the described procedure, we calculate the standard deviation of the lower bound estimate using the objective values calculated for each sample:

Sample	z_N^m	$z_N^m - \overline{z}_N$	$(z_N^m - \overline{z}_N)^2$
1	512,871	-73,950	54,690,462
2	547,378	27,111	735,044,276
3	551,181	30,914	955,718,676
4	510,364	-9,902	98,055,545
5	509,150	-11,116	123,572,125
6	517,043	-3,223	10,389,662
7	501,330	-18,936	358,583,457
8	520,173	-93	8,704
9	513,322	-6,944	48,223,302
10	519,851	-415	172,474

Table 4.2: Example: SAA Upper Bound Objective Values

Specifically, the variance of the estimator of the lower bound is calculated by the square root of the sum of the last column, divided by the amount of $(M-1)$:

$$\hat{\sigma}^2_{\overline{z}_N} = 16,277$$

With this information the confidence interval would be:

$$CI_{LB} = \pm 10,088$$
$$\overline{z}_N \in [510,178; 530,354]$$

For the upper bound, we choose the sample with the smallest objective function value (sample number 7). We use this solution's first stage variables values, $\hat{u}_{g,t}^{z_7}$ and $\hat{c}_{g,t}^{z_7}$ to solve a UCPSD for N' scenarios fixing the first stage variables $u_{g,t}$ and $c_{g,t}$ to $\hat{u}_{g,t}^{z_7}$ and $\hat{c}_{g,t}^{z_7}$. Given the relatively small S, we can solve $N' = S = 20$ scenarios. This is a relatively quick problem to solve as the constraints of the second stage are just a group of independant linear equations.

Scenario	z_N^m	$z_N^m - z_N$	$(z_N^m - z_N)^2$
1	695,575	135,251	18,292,916,299
2	764,161	203,837	41,549,403,961
3	648,861	88,536	7,838,702,636
4	514,444	-45,880	2,104,968,425
5	489,402	-70,923	5,030,016,432
6	542,053	-18,271	333,842,617
7	603,689	43,365	1,880,528,398
8	456,417	-103,907	10,796,721,840
9	589,023	28,699	823,613,838
10	526,600	-33,724	1,137,317,695
11	523,822	-36,502	1,332,425,581
12	548,675	-11,649	135,708,845
13	494,724	-65,600	4,303,414,387
14	520,975	-39,349	1,548,348,241
15	535,926	-24,398	595,259,424
16	561,514	1,189	1,414,669
17	492,206	-68,118	4,640,065,605
18	508,800	-51,524	2,654,696,579
19	574,775	14,451	208,837,901
20	614,842	54,518	2,972,158,095

Table 4.3: Example: SAA Lower Bound Objective Values

Following the formula (4.24) we calculate the estimate for the upper bound:

$$\hat{z}_{N'}(\hat{u}_{g,t}^{z_7}, \hat{c}_{g,t}^{z_7}) = 560,324$$

And as previously described in formula (4.26), we calculate the standard deviation of the upper bound estimate using the objective value found in each scenario:

$$\hat{\sigma}^2_{\hat{z}_{N'}(\hat{u}_{g,t}^{z_7}, \hat{c}_{g,t}^{z_7})} = 75,457$$

Following the formulation, the confidence interval would be:

$$CI_{UB} = \pm 46,768$$
$$\hat{z}_{N'}(\hat{u}_{g,t}^{z_7}, \hat{c}_{g,t}^{z_7}) \in [513,556; 607,091]$$

At last, we calculate the optimality gap, its variance and confidence interval for a 95% confidence (significance level $\alpha = 0.05$) with formulas (4.28), (4.29) and (4.30) respectively.

$$GAP_{M,N,N'} = 607,091 - 510,178 = 96,913$$

$$\sigma^2_{GAP_{M,N,N'}} = \frac{75,457}{\sqrt{20}} + \frac{16,277}{\sqrt{10}} = 16,873 - 5,147 = 22,020$$

$$GAP_{M,N,N'} \in \Big[73,931; 119,895\Big]$$

5 | Computational Study

In Section 5.1 we discussed the choice of data of the generators, and the adjustments we implemented to the original set of data in order to cope with the missing information. Then in Section 5.2, we described the demand and the scenario sampling process. Section 5.3 develops the procedure for the generation of different instances. Lastly, Section 5.4 describes the equipment and software used in the experiments.

5.1. Choice of Data

In order to make a correct decision on the choice of data it is important to underline the purpose of the model. In general, we would like to find a solution for a UCPSD given certain data. In our case, we are trying to demonstrate the performance of the SAA in efficiently solving the UCPSD when the number of scenarios is large. Therefore, the value of the specific solutions we found to our data is not relevant, outside the fact that we were able to find a solution. We will focus on the time it takes the algorithm to find a solution, and how close is this solution to the real one. As we are using a sampling method, we will come up with upper and lower bounds, so we will be analyzing the dispersion of these bounds and how centered they are with respect to the actual solution.

Due to the previously mentioned reasons, we have decided to use the same problem instances as the work of Magnus [21], "An updated version of the ‘IEEE RTS 24-Bus System for Electricity Market and Power System Operation Studies’ " [17], as we consider this a great opportunity to have some perspective. The data was designed for a network-constrained UCP, but we can omit the node distribution of generators and demand. These data set are not a representation of some real data, but are constructed with the purpose of being well suited for testing.

This data set has 12 generators and a system load for a 24 hs period. As mentioned before, each generator has start-up cost, production cost, minimum and maximum production, ramp-up and ramp-down times, minimum up and down time, initial state as well as initial production. However, this data set doesn’t provide commitment cost nor shedding cost.

For the first, we are gonna make an approximation. The same way it is expressed in Rimer [21], we will be assuming the commitment cost to be a 5% of the maximum production cost. This is an arbitrary decision, as the main goal of this paper is to understand the advantages of using SAA and not to be close to reality. The main idea, is to avoid generators to be unnecessary committed. Thus,

$$C_g^C = 0.05 P_g^{Max} C_g^P \tag{5.1}$$

For the shedding cost, we used the production cost as a reference. The objective is to avoid the shedding load if possible, so we want the shedding cost to be high enough to incentivize the model to overproduce (and sell it for free to the grid) as opposed to paying the cost of shedding. The main idea is to avoid infeasible solution in case of under production. Given that the biggest production cost value is 20.93 €/MB, a value of 200 €/MB is enough for our purpose. It could be argued that the shedding cost should be dependant on time. As mentioned the objective of this thesis is to evaluate the SAA and the exact value of the solution is only relevant for this endeavour.

$$L_t = 200 \text{ €/MB} \quad \forall t \in \mathcal{T} \tag{5.2}$$

5.2. Stochastic Demand

The original paper provides a fix discrete demand for each time period which will denoted with $D_t^{original}$. The data is provided in Appendix A. We would like to have a demand distribution in order to generate scenario set $\mathcal{S}$, and we want these scenarios to be the same for all experiments. First for each time period t, we will assume a continuous distribution of the demand centered at $D_t^{original}$ with a standard deviation of σ. Then we will randomly take the sample of scenarios from this continuous distribution using a random seed of value 1 to keep this scenario set constant between experiments. By doing so, we guarantee that the distribution is centered around the demand found in the data. The only problem would be to create negative values for the demand. This can be achieved by keeping the standard deviation relatively small. Notice that the samples for the SAA are chosen randomly, so the random seed is reset every time we take a sample from S.

5.3. Instance Generation

We took into consideration up to 24 generators. We created a new set of generators starting from the original set of 12 generators and properly multiplying the parameters $P_g^{Min}, P_g^{Max}, R_g^{Up}, R_g^{Down}, T_g^{Up}, T_g^{Down}, C_g^P, C_g^C, C_g^S$ and $pInit_g$ corresponding to each generator by a coefficient of 1.5 to obtain a new set of 12 generators with increased capabilities and associated costs. These 24 generators were used to solved the UCPSD problem for discrete demand.

The original data set was design to satisfy the demand previously mentioned, so by adding these new generators to the original ones we are not affecting the nature of the UCPSD problem. This is done with the soul purpose of giving information on the changes in the objective values as well as a deeper understanding of the limitations of our model and the computational system.

For solving the extensive formulation of the UCPSD we generated 15 instances. While for the SAA method, we considered 36 instances. For this, we considered two key parameters in the model: the **generators** and the **standard deviation** to generate the scenarios from the continuous probability demand function.

For the generators we considered different instances between the extensive UCPSD and the SAA method. For the first, we run the model for 8, 10, 12, 14 and 16 generators. While for the later, we took even sets from the complete set of generators, that's to say, we run the SAA for 2,4, ..., 24 generators.

Regarding the standard deviation σ, we previously mentioned that we need to keep it small enough to avoid problems related to generating negative values. After a few tests, we concluded that up to 15% would avoid this problem, so we choose three standard deviations of 5%,10% and 15% to run the tests. The standard deviation is expressed as a percentage given that the demand changes from one period to the other, meaning we need to adjust the σ_t value at each period of time to the value of the demand in that period in order to be representative. To normalize the coefficient, we took σ_t as a percentage of $D_t^{original}$ for each t.

$$\sigma_t = \sigma D_t^{original} \tag{5.3}$$

With σ_t it is possible to generate and sample the continuous function of demand. Just in case, any negative value resulting from the normalization of demand is considered as zero.

$$\mathcal{G}^{UCPSD} = \{8, 10, 12, 14, 16\} \qquad \mathcal{G}^{SAA} = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24\}$$
$$\sigma^{UCPSD} \in \{5\%, 10\%, 15\%\} \qquad \sigma^{SAA} \in \{5\%, 10\%, 15\%\}$$

To clarify, $\mathcal{G}^{UCPSD}$ and σ^{UCPSD} represents the different sets of generators and σ values respectively, used to run the experiments of the extensive formulation of the unit commitment problem with stochastic demand. While $\mathcal{G}^{SAA}$ and σ^{SAA} represents the ones used in the sample average approximation method.

5.4. Implementation Details

The implementation of the model was done in a DELL laptop with Microsoft Windows 11 pro, processor Intel(R) Core(TM) i7-10510U CPU @ 1.80 GHz 2.30 GHz equipped with 8 GB of **RAM memory**. **Python** version 3.10.10 was ran in Visual Basic Code and Jupyter Notebook. The **solver** of the model for both the UCPSD and the SAA was Gurobi Optimizer version 10.0.1 build v10.0.1rc0 (win64).

6 | Results

The numerical experimentation is performed in two parts. Firstly, we run the extensive formulation of the UCPSD for 15 instances, increasing the number of sample scenarios until the code was unable to find a solution for one instance in the stated time limit. In the second part, we decided on a comparatively large number of scenarios and run the code for the SAA method for 36 instances. Since the major motive of the present thesis is to analyze the performance of the SAA, we emphasize our display of data on the run times, the optimality gaps and confidence intervals.

6.1. Extensive Formulation UCPSD

We first analyse the extensive formulation of the UCPSD. Specifically, we solved the instances with $|\mathcal{S}| = 100$, 200, 500, 1000, 2000 and 3000 scenarios using 3 different standard deviations for the demand of $\sigma_{Demand} = 5\%$, 10%, 15% and 5 different generators set $|\mathcal{G}| =$ 8, 10, 12, 14, 16. The aggregate results are shown in Table 6.1, while the complete results can be found in Appendix B. The values in the last column represent the averages. We imposed to Gurobi a time limit of 3,600 seconds and a MIPGap of 0.5 to Gurobi. This gap tells the solver when to stoplooking for a solution.

	Scenarios $\|\mathcal{S}\|$						
$\|\mathcal{G}\|$	100	200	500	1,000	2,000	3,000	
8	2,255,251	2,244,964	2,247,735	2,242,712	2,241,018	2,238,803	2,245,080
10	774,508	768,297	770,480	769,474	768,160	768,296	769,869
12	578,110	575,914	577,520	577,040	577,451	576,748	577,131
14	576,903	566,844	568,134	567,796	568,491	570,065	569,705
16	576,084	567,080	568,003	567,597	567,519	-	569,257

Table 6.1: Objective Values: Total Costs

Table 6.1 presents the average results of experiments for all three values of σ_{Demand}. Thus, for each number of scenarios $|\mathcal{S}|$ three runs are carried out. The "-" symbol represents the incapacity of the solver to find a solution in the time restriction imposed.

Results in Table 6.1 indicate that the objective value of the solution is relatively unchanged with an increasing amount of scenarios. Moreover, we observe that the extensive formulation for less than 12 generators obtains a larger total cost. On the other hand, increasing over 12 generators does not provide a considerable decrease in the total costs. This makes sense as the data we are using [17] is designed for 12 generators.

We calculated the standard deviation between the objective function values for different scenarios with the same $|\mathcal{G}|$ and σ_{Demand} and present the results in Table 6.2

	σ_{Demand}			
$\|\mathcal{G}\|$	5%	10%	15%	
8	2,500	5,516	8,068	5,362
10	1,402	2,304	3,122	2,276
12	1,795	1,071	1,170	1,346
14	5,307	11,268	1,736	6,104
16	522	10,682	521	3,909

Table 6.2: Standard deviation of the objective values z along all scenarios for the same $|\mathcal{G}|$ and σ_{Demand}

Results in Table 6.2 confirm that the objective values of the solution are relatively unchanged with the number of scenarios.

	σ_{Demand}=5%		σ_{Demand}=10%		σ_{Demand}=15%	
$\|\mathcal{G}\|$	z	Time	z	Time	z	Time
8	2,198,842	20	2,232,398	32	2,304,002	26
10	675,912	132	758,032	163	875,664	275
12	561,571	515	573,370	462	596,451	346
14	543,587	1,231	562,812	826	602,717	614

Table 6.3: Average objective values and run times variation with respect to the σ values for all scenarios

Table 6.3 presents the average results of experiments for all scenarios for all six values of $|\mathcal{S}|$ where the number of generators $|\mathcal{G}|$ ranges from 8 to 16. Results in this table present the variation in the objective values and run times with respect to the σ_{Demand} values. Regarding the total cost, we note a slight tendency to increase with increasing standard deviation σ_{Demand}. As demand deviates from the mean, new generators needs to be turned on in order to satisfy demand.

On the other hand, the run times have a more unpredictable behaviour than the standard deviation. For 14 generators, the total cost decreases with increasing σ_{Demand} while for 10, it increases for increasing σ_{Demand}

	Scenarios $|\mathcal{S}|$					
$|\mathcal{G}|$	100	200	500	1000	2000	3000
8	1	2	9	17	48	78
10	3	5	36	92	296	708
12	8	21	141	182	855	1,439
14	17	23	206	404	1,896	2,796
16	37	35	238	526	2,709	-
	13	17	126	244	1161	1255

Table 6.4: Average Run Times in seconds

Table 6.4 results show the average run times corresponding to the entries in Table 6.1. We note that run times significantly grow with the number of scenarios. Again, the "-" indicates that the solver was unable to find a solution in less than 60 minutes, as imposed to the solver. We conclude that the problem becomes impractical to solve for large instances. This justifies the need for developing the SAA method.

6.2. Sample Average Approximation

We run the SAA method for 36 instances with $|\mathcal{S}| = 10,000$ scenarios for $M = 30$ samples of size $N = 50, 100, 150, 200$ for the lower bound and $N' = 40N$ for the upper bound, i.e., $N' = 2000, 4000, 6000, 8000$ respectively. For each run we computed the upper and lower bound values, as well as their corresponding run times, confidence intervals and variance as described in Section 4.1.

	Sample Size N				
$|\mathcal{G}|$	50	100	50	200	
2	9,877,827	9,879,316	9,873,277	9,879,625	9,877,511
4	5,265,819	5,263,574	5,269,542	5,269,039	5,266,994
6	4,333,204	4,338,234	4,337,726	4,336,931	4,336,524
8	2,239,434	2,240,058	2,244,383	2,239,400	2,240,819
10	692,032	693,231	692,613	691,934	692,452
12	501,072	500,771	500,992	500,503	500,835
14	496,152	492,778	493,593	492,287	493,703
16	494,337	492,075	491,471	493,674	492,889
18	478,299	478,591	477,746	476,916	477,888
20	418,745	419,336	419,379	419,710	419,292
22	281,375	281,730	281,637	281,428	281,542
24	281,765	281,443	281,724	281,705	281,659

Table 6.5: Average Lower Bound values for 50, 100, 150 and 200 scenarios sample size for σ values of 5%,10% and 15%

We will present aggregated results and average values relevant for our discussion. We observe in Table 6.5 a reduction in total cost as we increase the number of generators in all instances, a similar behaviour as appreciated in Table 6.1. Note the insignificant variation in the objective values with respect to the sample size. We showed only the lower bounds results in this table, but the same trend can be appreciated for the upper bound results.

$\|\mathcal{G}\|$	LB	UB	Run Time	$\frac{UB-LB}{LB}$	σ_{GAP}
2	9,877,511 ± 7,246	9,878,325 ± 78,237	3	0.9%	70,991
4	5,266,994 ± 7,952	5,270,255 ± 77,942	8	1.7%	69,990
6	4,336,524 ± 7,694	4,335,391 ± 77,141	15	1.9%	69,447
8	2,240,819 ± 6,792	2,242,981 ± 67,191	48	3.4%	60,398
10	692,452 ± 2,827	692,279 ± 28,088	96	4.2%	25,262
12	500,835 ± 1,010	498,989 ± 7,809	336	1.4%	6,799
14	493,703 ± 2,125	490,084 ± 7,170	550	1.1%	5,045
16	492,889 ± 2,078	490,522 ± 7,202	1,097	1.4%	5,124
18	477,888 ± 2,083	475,090 ± 7,141	1,246	1.3%	5,058
20	419,292 ± 1,195	417,706 ± 6,460	1,152	1.4%	5,265
22	281,542 ± 877	277,748 ± 5,592	902	0.9%	4,716
24	281,659 ± 1,029	277,434 ± 6,390	1,341	1.1%	5,361

Table 6.6: Average results of the SAA for σ_{Demand} values of 5%,10%, 15% and sample size N=50,100,150,200

We can appreciate in Table 6.6 that the run times are much smaller than those of Table 6.4 for 3,000 scenarios for the same number of generators $\mathcal{G}$, regardless of the considerable difference in the scenarios considered. In this table the Run Time considers both the time to find both the upper and lower bound. In general, the time taken to find the upper bound, i.e., the time taken to evaluate the N' sample, is insignificant with respect to the time taken to run the $M = 30$ samples of N size and find the lower bound. For this reason we indicated in Table 6.6 the total time. The specific time it took the experiment to find the lower and upper bound respectively can be found in the Appendix B. Note that the total cost in this case kept decreasing for instances with more than 12 generators.

Note that for 12 generators, the SAA method took an average of 5 minutes for each instance to find the upper and lower bounds for a $|\mathcal{S}| = 10,000$ scenarios. While on Table 6.4, we can appreciate that it took the solver 20 minutes to find an exact solution for 12 generators and $|\mathcal{S}| = 3,000$. This shows a considerable reduction in the time required to find a solution.

Moreover, the confidence intervals of both the lower and the upper bound are considerable small. We can appreciate a consistent decrease on the interval between 8 generators and 10 generators. At 10 generators the data seems to stabilize, and the minimum confidence interval is reached for 22 generators.

Note that the optimality gaps doesn't go beyond 2% in almost all instances and the variance of the GAP, i.e., σ_{GAP}, stays below 7,000 from 12 generators onwards, which represents 1% and 2% of the lower bound value, providing high consistency to the results.

7 | Conclusions and future developments

In this thesis we introduced the Unit Commitment Problem with Stochastic Demand and described a mathematical for its extensive formulation that minimizes total costs. We the proposed the Sample Average Approximation methodology to solve instances with large number of scenarios and studied its efficiency. The computational experiments showed that by using the SAA method we can achive a considerable decrease in the time needed to find a solution with small optimality gaps. From testing this methodology it became clear how taking advantage of dividing big problems into smaller ones can beneficial without losing quality of result. The SAA proved to be relatively simple and effective; we believe it can be very usefull in solving other scheduling problem with the same efficiency.

To conclude we believe that the thesis paves the way to several avenues of future research. We note that while we showed that the SAA provides high quality solutions, the run times are directly proportional to the Sample size N. Further research could be done in trying to find the optimal number of samples M and sample size N for the number of scenarios $|\mathcal{S}|$.

We described the UCPSD, but uncertainty can be found in many parameters (e.g. production). As mentioned, nowadays the trend is changing and so are the formulations of the UCPSD. In this thesis, we wanted to evaluate the performance of the SAA in solving the UCPSD, but our work could be further developed taking into account other random parameters or develop the formulation to make the model more realistic.

Bibliography

[1] H. Abdi. Profit-based unit commitment problem: A review of models, methods, challenges, and future directions. *Renewable and Sustainable Energy Reviews*, 138: 110504, 11 2020.

[2] S. Ahmed and A. Shapiro. The sample average approximation method for stochastic programs with integer recourse. *Science*, 2002.

[3] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer New York, NY, 2011.

[4] T. H. de Mello and G. Bayraksan. Monte carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, 19, 2014. ISSN 18767354.

[5] J. Dupacová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming. *Mathematical Programming*, 95:493–511, 2003.

[6] Z. Geng, A. Conejo, Q. Chen, and C. Kang. Power generation scheduling considering stochastic emission limits. *International Journal of Electrical Power & Energy Systems*, 95:374–383, 02 2018.

[7] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer New York, NY, 2003.

[8] Y. Huang, Q. P. Zheng, and J. Wang. Two-stage stochastic unit commitment model including non-generation resources with conditional value-at-risk constraints. *Electric Power Systems Research*, 2014.

[9] M. Håberg. Fundamentals and recent developments in stochastic unit commitment. *International Journal of Electrical Power & Energy Systems*, 109:38–48, 2019.

[10] A. J. Kleywegt, A. Shapiro, and T. H. de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.*, 12:479–502, 2002.

[11] D. Kothari and A. Ahmad. An expert system approach to the unit commitment problem. *Energy Conversion and Management*, pages 257–261, 1995.

[12] W. Mak, D. Morton, and R. Wood. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1):47–56, Feb. 1999.

[13] R. Mallipeddi and P. Suganthan. Unit commitment - a survey and comparison of conventional and nature inspired algorithms. *Int. J. of Bio-Inspired Computation*, 6: 71 – 90, 01 2014.

[14] L. Montero, A. Bello, and J. Reneses. A review on the unit commitment problem: Approaches, techniques, and resolution methods. *Energies*, 15(4), 2022.

[15] V. Norkin. Global optimization of probabilities by the stochastic branch and bound method. In K. Marti and P. Kall, editors, *Stochastic Programming Methods and Technical Applications*, pages 186–201. Springer Berlin Heidelberg, 1998.

[16] W. Ongsakul and N. Petcharaks. Unit commitment by enhanced adaptive lagrangian relaxation. *Power Systems, IEEE Transactions on*, 19:620 – 628, 03 2004.

[17] C. Ordoudis, P. Pinson, J. Morales González, and M. Zugno. *An Updated Version of the IEEE RTS 24-Bus System for Electricity Market and Power System Operation Studies.* Technical University of Denmark, 2016.

[18] N. Padhy. Unit commitment-a bibliographical survey. *IEEE Transactions on Power Systems*, 19(2):1196–1205, 2004.

[19] D. Palis and S. Palis. Efficient unit commitment - a modified branch-and-bound approach. In *2016 IEEE Region 10 Conference (TENCON)*, pages 267–271, 2016.

[20] K. Ponomareva, D. Roman, and P. Date. An algorithm for moment-matching scenario generation with application to financial portfolio optimisation. *European Journal of Operational Research*, 240(3):678–687, 2015.

[21] M. Rimer. Master thesis in mathematics-economics. Master's thesis, Department of Mathematical Sciences, 2022.

[22] L. A. C. Roque. *Optimization Methods for the Unit Commitment Problem in Electric Power Systems.* PhD thesis, School of Mathematics, 2014.

[23] A. Sayed, M. Ebeed, Z. Ali, A. Bedair, A. Abdel-Rahman, M. Ahmed, S. Abdel Aleem, A. El-Shahat, and M. Rihan. A hybrid optimization algorithm for solving

of the unit commitment problem considering uncertainty of the load demand. *Energies*, 2021.

[24] P. Seljom and A. Tomasgard. Sample average approximation and stability tests applied to energy system design. *Energy Systems*, 12, 02 2021.

[25] T. Shiina. Unit commitment problem with stochastic demand. *Journal of Computations & Modelling*, 2012.

[26] T. Shiina and J. R. Birge. Stochastic unit commitment problem. *International Transactions in Operational Research*, 11, 2004. ISSN 14753995.

[27] S. Takriti, J. Birge, and E. Long. A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems*, 11(3):1497–1508, 1996.

[28] B. Verweij, S. Ahmed, A. Kleywegt, G. Nemhauser, and A. Shapiro. The sample average approximation method applied to stochastic routing problems: A computational study. *Computational Optimization and Applications*, 24:289–333, 02 2003.

[29] Q. Wang, J. Wang, and Y. Guan. Stochastic unit commitment with uncertain demand response. *IEEE Transactions on Power Systems*, 2013.

[30] Y.-G. Xie and H.-D. Chiang. A novel solution methodology for solving large-scale thermal unit commitment problems. *Electric Power Components and Systems*, 38 (14):1615–1634, 2010.

[31] P. Xiong and P. Jirutitijaroen. A stochastic optimization formulation of unit commitment with reliability constraints. *IEEE Transactions on Smart Grid*, 4(4):2200–2208, 2013.

[32] W. Zhi, Z. Pingliang, Z. Xiao-Ping, and Z. Qinyong. A solution to the chance-constrained two-stage stochastic program for unit commitment with wind energy integration. *IEEE Transactions on Power Systems*, 31(6):4185–4196, 2016.

[33] J. Zou, S. Ahmed, and X. A. Sun. Multistage stochastic unit commitment using stochastic dual dynamic integer programming. *IEEE Transactions on Power Systems*, 2019.

A | Appendix A - Data

Hour	**System Demand** (MW)	**Hour**	**System Demand** (MW)
1	1775.835	13	2517.975
2	1669.815	14	2517.975
3	1590.3	15	2464.965
4	1563.795	16	2464.965
5	1563.795	17	2623.995
6	1590.3	18	2650.5
7	1961.37	19	2650.5
8	2279.43	20	2544.48
9	2517.975	21	2411.955
10	2544.48	22	2199.915
11	2544.48	23	1934.865
12	2517.975	24	1669.815

Table A.1: System demand

Unit	P_g^{Min}	P_g^{Max}	R_g^{Up}	R_g^{Down}	T_g^{Up}	T_g^{Down}	C_g^P	C_g^C	C_g^S	$pInit_g$	$uInit_g$
1	30.4	152	120	120	8	4	13.32	101	1430.4	76	1
2	30.4	152	120	120	8	4	13.32	101	1430.4	76	1
3	75	350	350	350	8	8	20.7	362	1725	0	0
4	206.85	591	240	240	12	10	20.93	618	3056.7	0	0
5	12	60	60	60	4	2	26.11	78	437	0	0
6	54.24	155	155	155	8	8	10.52	82	312	0	0
7	54.24	155	155	155	8	8	10.52	82	312	124	1
8	100	400	280	280	1	1	6.02	120	0	240	1
9	100	400	280	280	1	1	5.47	109	0	240	1
10	300	300	300	300	0	0	0	0	0	240	1
11	108.5	310	180	180	8	8	10.52	163	624	248	1
12	140	350	240	240	8	8	10.89	191	2298	280	1

Table A.2: Generator's data

B | Appendix B - Results UCPSD

$\lvert\mathcal{G}\rvert$	$\lvert\mathcal{S}\rvert$	σ_{Demand}	**Time**	**Z**
8	100	5%	1	2,203,440
10	100	5%	2	678,569
12	100	5%	8	564,418
14	100	5%	18	540,570
16	100	5%	36	539,908
8	100	10%	1	2,242,761
10	100	10%	3	762,676
12	100	10%	9	572,360
14	100	10%	16	587,885
16	100	10%	38	586,089
8	100	15%	1	2,319,552
10	100	15%	3	882,280
12	100	15%	7	597,553
14	100	15%	17	602,254
16	100	15%	36	602,254
8	200	5%	2	2,198,768
10	200	5%	4	674,408
12	200	5%	20	559,009
14	200	5%	31	539,720
16	200	5%	38	539,720
8	200	10%	2	2,232,878
10	200	10%	5	755,916
12	200	10%	23	571,754
14	200	10%	21	558,197
16	200	10%	34	558,905
8	200	15%	2	2,303,245
10	200	15%	5	874,568
12	200	15%	19	596,980
14	200	15%	18	602,614

Table B.1: Extensive formulation of the Unit Commitment Problem with Sotchastic Demand solved with Cplex

$\lvert\mathcal{G}\rvert$	$\lvert\mathcal{S}\rvert$	$\sigma_{\textbf{Demand}}$	**Time**	**Z**
16	200	15%	32	602,614
8	500	5%	8	2,200,298
10	500	5%	34	676,866
12	500	5%	196	559,659
14	500	5%	241	540,761
16	500	5%	249	540,943
8	500	10%	10	2,235,065
10	500	10%	36	758,889
12	500	10%	134	575,035
14	500	10%	263	559,473
16	500	10%	246	559,473
8	500	15%	8	2,307,842
10	500	15%	39	875,684
12	500	15%	92	597,867
14	500	15%	114	604,168
16	500	15%	220	603,594
8	1000	5%	17	2,197,701
10	1000	5%	88	675,366
12	1000	5%	223	561,830
14	1000	5%	460	539,526
16	1000	5%	540	540,850
8	1000	10%	16	2,229,627
10	1000	10%	78	757,813
12	1000	10%	163	574,014
14	1000	10%	454	558,290
16	1000	10%	493	559,661
8	1000	15%	18	2,300,809
10	1000	15%	110	875,243
12	1000	15%	158	595,276
14	1000	15%	299	605,571
16	1000	15%	545	602,281
8	2000	5%	40	2,197,201
10	2000	5%	219	675,291
12	2000	5%	1079	562,335
14	2000	5%	3315	546,894
16	2000	5%	4443	540,817
8	2000	10%	65	2,228,151
10	2000	10%	276	756,418
12	2000	10%	722	573,613
14	2000	10%	1568	557,178

Table B.1: Extensive formulation of the Unit Commitment Problem with Stochastic Demand solved with Cplex (continue)

$\lvert\mathcal{G}\rvert$	$\lvert\mathcal{S}\rvert$	**sigma_D**	**Time**	**Z**
16	2000	10%	2174	559,524
8	2000	15%	39	2,297,702
10	2000	15%	394	872,772
12	2000	15%	764	596,404
14	2000	15%	804	601,401
16	2000	15%	1508	602,216
8	3000	5%	54	2,195,641
10	3000	5%	444	674,973
12	3000	5%	1566	562,175
14	3000	5%	3322	554,052
16	3000	5%	-	-
8	3000	10%	96	2,225,908
10	3000	10%	578	756,481
12	3000	10%	1717	573,444
14	3000	10%	2635	555,849
16	3000	10%	-	-
8	3000	15%	85	2,294,859
10	3000	15%	1101	873,435
12	3000	15%	1032	594,624
14	3000	15%	2431	600,293
16	3000	15%	-	-

Table B.1: Extensive formulation of the Unit Commitment Problem with Sotchastic Demand solved with Cplex (continue)

C | Appendix C - Results SAA

The following tables presents the results of the experiments with the Sample Average Approximation method using $M = 30$ samples per run for $|\mathcal{G}| \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24\}$ with $\sigma_{Demand} \in \{0.05, 0.1, 0.15\}$ and $N \in \{50, 100, 150, 200\}$

$\|\mathcal{G}\|$	M	σ_D	N	Time (LB)	Time (UB)	LB	UB	σ_{LB}	σ_{UB}	GAP	σ_{GAP}
2	30	5%	50	1.1	0.3	9,877,015 ± 6,702	9,877,334 ± 39,367	18,635	109,455	46,388 ± 90,754	46,069
2	30	5%	100	1.8	0.6	9,878,292 ± 4,106	9,878,134 ± 38,613	11,417	107,358	42,561 ± 84,154	42,719
2	30	5%	150	2.5	0.9	9,876,942 ± 3,766	9,877,371 ± 38,938	10,472	108,263	43,133 ± 84,126	42,704
2	30	5%	200	3.4	1.2	9,876,838 ± 2,758	9,876,794 ± 39,187	7,668	108,955	41,901 ± 82,629	41,945
2	30	10%	50	0.7	0.2	9,877,244 ± 9,887	9,878,253 ± 79,627	27,489	221,395	90,523 ± 176,338	89,514
2	30	10%	100	1.5	0.5	9,874,748 ± 6,831	9,876,005 ± 78,308	18,992	217,727	86,396 ± 167,719	85,139
2	30	10%	150	2.4	1.1	9,873,049 ± 6,762	9,880,459 ± 77,542	18,801	215,597	91,714 ± 166,075	84,304
2	30	10%	200	3.8	1.5	9,881,245 ± 6,413	9,880,146 ± 77,956	17,831	216,749	83,271 ± 166,204	84,370
2	30	15%	50	0.8	0.4	9,879,223 ± 11,841	9,869,346 ± 115,609	32,922	321,439	117,573 ± 251,071	127,450
2	30	15%	100	2.0	0.7	9,884,909 ± 11,892	9,883,069 ± 118,805	33,063	330,323	128,856 ± 257,465	130,696
2	30	15%	150	2.9	1.0	9,869,839 ± 8,365	9,882,550 ± 117,585	23,258	326,932	138,661 ± 248,115	125,950
2	30	15%	200	3.3	1.3	9,880,793 ± 7,634	9,880,444 ± 117,309	21,226	326,164	124,594 ± 246,131	124,943
4	30	5%	50	2.8	0.5	5,262,084 ± 6,545	5,263,529 ± 38,710	18,198	107,628	46,700 ± 89,150	45,255
4	30	5%	100	3.0	1.2	5,263,008 ± 4,085	5,263,582 ± 38,591	11,359	107,299	43,251 ± 84,071	42,677
4	30	5%	150	4.9	3.1	5,264,505 ± 3,511	5,264,312 ± 39,333	9,763	109,362	42,652 ± 84,402	42,845
4	30	5%	200	6.4	4.4	5,264,019 ± 2,906	5,264,603 ± 39,052	8,079	108,581	42,542 ± 82,655	41,958
4	30	10%	50	2.0	1.0	5,269,006 ± 12,175	5,269,516 ± 77,283	33,851	214,877	89,968 ± 176,228	89,458
4	30	10%	100	3.5	2.0	5,267,151 ± 8,090	5,269,149 ± 78,891	22,494	219,348	88,979 ± 171,349	86,981
4	30	10%	150	5.3	3.3	5,271,506 ± 6,945	5,264,296 ± 77,349	19,309	215,059	77,083 ± 166,054	84,293
4	30	10%	200	7.1	4.5	5,269,894 ± 4,827	5,266,116 ± 78,414	13,421	218,022	79,463 ± 163,981	83,241
4	30	15%	50	2.7	1.2	5,266,368 ± 20,499	5,287,243 ± 118,467	56,995	329,383	159,840 ± 273,755	138,965
4	30	15%	100	4.7	2.5	5,260,562 ± 10,093	5,279,927 ± 115,789	28,063	321,939	145,247 ± 247,982	125,882
4	30	15%	150	7.0	4.0	5,272,616 ± 9,067	5,275,457 ± 116,398	25,210	323,631	128,306 ± 247,159	125,465
4	30	15%	200	9.8	11.0	5,273,204 ± 6,680	5,275,326 ± 117,025	18,573	325,375	125,827 ± 243,693	123,705

Table C.1: Sample Average Approximation Method

$\lvert\mathcal{G}\rvert$	M	σ_D	N	Time (LB)	Time (UB)	LB	UB	σ_{LB}	σ_{UB}	GAP	σ_{GAP}
6	30	5%	50	3.4	1.3	4,310,858 ± 6,416	4,317,001 ± 39,470	17,840	109,742	52,029 ± 90,394	45,886
6	30	5%	100	5.2	2.7	4,318,778 ± 4,309	4,314,028 ± 39,357	11,981	109,428	38,916 ± 86,020	43,666
6	30	5%	150	8.0	4.8	4,313,146 ± 3,390	4,314,858 ± 38,931	9,425	108,244	44,033 ± 83,370	42,321
6	30	5%	200	11	8.4	4,316,293 ± 2,886	4,314,532 ± 38,843	8,025	108,000	39,969 ± 82,206	41,730
6	30	10%	50	3.7	1.6	4,333,929 ± 11,797	4,330,995 ± 78,512	32,799	218,294	87,375 ± 177,903	90,309
6	30	10%	100	7.6	3.3	4,332,477 ± 7,427	4,331,255 ± 76,644	20,649	213,101	82,849 ± 165,616	84,071
6	30	10%	150	11	5.1	4,331,349 ± 6,458	4,329,138 ± 77,646	17,956	215,885	81,893 ± 165,680	84,104
6	30	10%	200	14	7.8	4,333,546 ± 5,129	4,332,603 ± 77,363	14,261	215,098	81,549 ± 162,505	82,492
6	30	15%	50	4.1	1.9	4,354,825 ± 17,741	4,358,313 ± 115,938	49,328	322,352	137,167 ± 263,341	133,679
6	30	15%	100	8.4	3.7	4,363,446 ± 11,883	4,357,342 ± 113,636	33,040	315,951	119,415 ± 247,266	125,519
6	30	15%	150	12	5.7	4,368,683 ± 7,700	4,360,894 ± 114,863	21,410	319,364	114,774 ± 241,444	122,563
6	30	15%	200	30	10	4,360,955 ± 7,190	4,363,728 ± 114,492	19,990	318,332	124,455 ± 239,707	121,682
8	30	5%	50	10	4.0	2,196,027 ± 4,367	2,198,773 ± 35,624	12,143	99,049	42,738 ± 78,781	39,992
8	30	5%	100	29	9.2	2,198,344 ± 3,696	2,198,066 ± 34,676	10,275	96,412	38,093 ± 75,589	38,371
8	30	5%	150	47	27	2,198,919 ± 2,854	2,196,829 ± 34,812	7,935	96,790	35,576 ± 74,199	37,666
8	30	5%	200	60	19	2,197,242 ± 2,070	2,196,237 ± 35,105	5,755	97,605	36,170 ± 73,232	37,175
8	30	10%	50	8.2	3.4	2,223,340 ± 10,004	2,238,832 ± 68,240	27,816	189,735	93,737 ± 154,138	78,245
8	30	10%	100	26	7.8	2,226,679 ± 7,227	2,229,510 ± 69,479	20,093	193,178	79,536 ± 151,106	76,705
8	30	10%	150	43	12	2,233,933 ± 4,456	2,227,662 ± 68,170	12,389	189,540	66,355 ± 143,070	72,626
8	30	10%	200	68	20	2,226,073 ± 5,337	2,228,217 ± 68,602	14,838	190,739	76,082 ± 145,655	73,938
8	30	15%	50	8.8	3.7	2,298,934 ± 12,651	2,302,844 ± 97,919	35,176	272,252	114,480 ± 217,818	110,570
8	30	15%	100	26	7.6	2,295,152 ± 11,139	2,300,735 ± 98,246	30,971	273,163	114,968 ± 215,484	109,385
8	30	15%	150	41	17	2,300,298 ± 9,740	2,301,391 ± 97,439	27,081	270,919	108,272 ± 211,138	107,179
8	30	15%	200	55	21	2,294,886 ± 7,968	2,296,678 ± 97,974	22,154	272,407	107,734 ± 208,701	105,942

Table C.1: Sample Average Approximation Method (continue)

$\lvert\mathcal{G}\rvert$	M	$\sigma_{\mathbf{D}}$	N	Time (LB)	Time (UB)	LB	UB	$\sigma_{\mathbf{LB}}$	$\sigma_{\mathbf{UB}}$	GAP	$\sigma_{\mathbf{GAP}}$
10	30	5%	50	18	5.4	600,796 ± 1,959	599,215 ± 9,253	5,446	25,728	9,631 ± 22,087	11,212
10	30	5%	100	58	11	599,546 ± 1,005	599,259 ± 9,203	2,793	25,588	9,921 ± 20,108	10,208
10	30	5%	150	92	20	599,763 ± 768	598,980 ± 9,379	2,134	26,077	9,363 ± 19,988	10,146
10	30	5%	200	135	28	598,995 ± 557	599,501 ± 9,347	1,548	25,988	10,410 ± 19,510	9,904
10	30	10%	50	16	4.9	679,272 ± 3,309	680,465 ± 27,331	9,201	75,992	31,834 ± 60,361	30,641
10	30	10%	100	60	10	682,545 ± 2,315	680,625 ± 27,263	6,436	75,801	27,658 ± 58,266	29,578
10	30	10%	150	98	20	678,852 ± 2,316	680,040 ± 27,399	6,439	76,180	30,903 ± 58,537	29,715
10	30	10%	200	144	29	678,835 ± 1,627	680,728 ± 27,455	4,523	76,336	30,975 ± 57,290	29,082
10	30	15%	50	19	6.0	796,028 ± 7,835	797,357 ± 47,725	21,785	132,695	56,890 ± 109,452	55,561
10	30	15%	100	58	11	797,601 ± 5,026	797,623 ± 47,854	13,974	133,054	52,902 ± 104,172	52,880
10	30	15%	150	98	21	799,225 ± 3,853	796,080 ± 47,147	10,714	131,087	47,855 ± 100,468	51,000
10	30	15%	200	165	29	797,971 ± 3,352	797,471 ± 47,701	9,319	132,628	50,553 ± 100,572	51,053
12	30	5%	50	70	4.0	486,064 ± 1,463	479,985 ± 3,308	4,068	9,197	-1,308 ± 9,398	4,771
12	30	5%	100	185	12	486,212 ± 593	483,886 ± 3,579	1,649	9,950	1,846 ± 8,218	4,172
12	30	5%	150	355	22	485,610 ± 601	483,291 ± 4,058	1,672	11,284	2,341 ± 9,180	4,660
12	30	5%	200	565	33	485,589 ± 518	483,239 ± 3,975	1,440	11,053	2,143 ± 8,851	4,4937
12	30	10%	50	71	5.9	498,118 ± 1,136	494,199 ± 6,147	3,159	17,092	3,365 ± 14,348	7,284
12	30	10%	100	227	14	497,665 ± 767	495,573 ± 6,583	2,133	18,304	5,258 ± 14,480	7,350
12	30	10%	150	428	18	498,256 ± 719	494,965 ± 6,435	1,999	17,892	3,863 ± 14,093	7,154
12	30	10%	200	699	33	497,646 ± 615	495,063 ± 6,542	1,711	18,188	4,574 ± 14,099	7,157
12	30	15%	50	57	6.5	519,035 ± 1,990	518,247 ± 12,723	5,534	35,374	13,925 ± 28,984	14,713
12	30	15%	100	167	13	518,437 ± 1,706	523,799 ± 15,814	4,742	43,969	22,881 ± 34,513	17,519
12	30	15%	150	344	26	519,110 ± 1,311	519,025 ± 12,567	3,644	34,941	13,793 ± 27,338	13,878
12	30	15%	200	618	52	518,273 ± 700	516,594 ± 11,980	1,947	33,310	11,002 ± 24,980	12,681

Table C.1: Sample Average Approximation Method (continue)

$\lvert\mathcal{G}\rvert$	M	$\sigma_{\mathbf{D}}$	N	Time (LB)	Time (UB)	LB	UB	$\sigma_{\mathbf{LB}}$	$\sigma_{\mathbf{UB}}$	GAP	$\sigma_{\mathbf{GAP}}$
14	30	5%	50	146	5.4	466,385 ± 1,436	463,441 ± 3,528	3,994	9,808	2,020 ± 9,779	4,964
14	30	5%	100	418	14	464,880 ± 1,052	463,600 ± 3,632	2,925	10,098	3,404 ± 9,227	4,684
14	30	5%	150	915	25	465,113 ± 667	464,972 ± 3,318	1,854	9,225	3,844 ± 7,850	3,985
14	30	5%	200	788	28	465,333 ± 860	464,276 ± 3,303	2,392	9,184	3,106 ± 8,202	4,163
14	30	10%	50	136	6.7	496,489 ± 5,473	483,413 ± 7,329	15,218	20,377	-274 ± 25,220	12,802
14	30	10%	100	376	13	487,778 ± 4,260	482,076 ± 7,103	11,845	19,749	5,661 ± 22,385	11,363
14	30	10%	150	873	25	489,549 ± 4,414	482,246 ± 7,192	12,274	19,997	4,304 ± 22,865	11,607
14	30	10%	200	659	35	484,767 ± 2,770	481,663 ± 7,534	7,703	20,948	7,201 ± 20,300	10,305
14	30	15%	50	125	8.6	525,583 ± 1,414	526,202 ± 11,896	3,931	33,075	13,929 ± 26,219	13,310
14	30	15%	100	409	17	525,677 ± 1,245	523,069 ± 10,714	3,462	29,788	9,351 ± 23,558	11,959
14	30	15%	150	877	33	526,118 ± 1,079	522,465 ± 10,617	3,001	29,518	8,043 ± 23,040	11,696
14	30	15%	200	621	44	526,762 ± 828	523,589 ± 9,872	2,301	27,449	7,527 ± 21,078	10,700
16	30	5%	50	188	6.2	466,425 ± 1,366	464,036 ± 3,549	3,798	9,868	2,526 ± 9,683	4,915
16	30	5%	100	616	18	464,640 ± 787	463,484 ± 3,272	2,188	9,098	2,903 ± 7,996	4,059
16	30	5%	150	1,528	22	465,211 ± 897	464,216 ± 3,360	2,495	9,342	3,262 ± 8,387	4,257
16	30	5%	200	1,178	54	466,236 ± 1,275	464,088 ± 3,305	3,545	9,189	2,432 ± 9,022	4,580
16	30	10%	50	286	7.7	489,870 ± 4,711	484,019 ± 8,237	13,098	22,902	7,097 ± 25,507	12,948
16	30	10%	100	790	20	486,764 ± 4,219	482,123 ± 7,414	11,730	20,613	6,992 ± 22,916	11,633
16	30	10%	150	2,297	30	483,480 ± 2,695	481,508 ± 7,531	7,492	20,940	8,254 ± 20,145	10,226
16	30	10%	200	1,123	42	488,555 ± 4,289	482,975 ± 7,440	11,925	20,686	6,149 ± 23,105	11,729
16	30	15%	50	208	8.5	526,717 ± 1,718	527,054 ± 10,300	4,776	28,638	12,355 ± 23,674	12,018
16	30	15%	100	1,124	24	524,821 ± 1,276	521,931 ± 10,867	3,548	30,214	9,253 ± 23,921	12,143
16	30	15%	150	2,340	41	525,722 ± 725	525,403 ± 10,623	2,017	29,537	11,030 ± 22,357	11,349
16	30	15%	200	1,121	96	526,232 ± 978	525,430 ± 10,520	2,720	29,251	10,697 ± 22,652	11,499

Table C.1: Sample Average Approximation Method (continue)

$\|\mathcal{G}\|$	M	$\sigma_{\mathbf{D}}$	N	Time (LB)	Time (UB)	LB	UB	$\sigma_{\mathbf{LB}}$	$\sigma_{\mathbf{UB}}$	GAP	$\sigma_{\mathbf{GAP}}$
18	30	5%	50	287	8.6	455,517 ± 641	455,389 ± 3,357	1,781	9,335	3,870 ± 7,876	3,998
18	30	5%	100	1,152	24	456,103 ± 634	455,797 ± 3,459	1,764	9,618	3,788 ± 8,064	4,094
18	30	5%	150	1,997	33	455,658 ± 290	455,366 ± 3,190	805	8,870	3,188 ± 6,855	3,480
18	30	5%	200	2,703	46	455,449 ± 291	455,219 ± 3,485	809	9,691	3,546 ± 7,439	3,776
18	30	10%	50	354	8.6	477,558 ± 1,965	474,035 ± 7,569	5,464	21,044	6,011 ± 18,781	9,534
18	30	10%	100	1,467	25	477,680 ± 1,617	473,745 ± 6,736	4,497	18,729	4,419 ± 16,456	8,354
18	30	10%	150	1,235	43	477,027 ± 1,424	473,315 ± 6,797	3,959	18,897	4,508 ± 16,194	8,220
18	30	10%	200	1,536	60	476,877 ± 1,599	473,067 ± 6,683	4,447	18,580	4,472 ± 16,315	8,282
18	30	15%	50	331	11	501,823 ± 4,919	497,153 ± 12,250	13,677	34,060	12,499 ± 33,822	17,169
18	30	15%	100	1,218	29	501,990 ± 4,668	495,457 ± 10,679	12,979	29,692	8,814 ± 30,233	15,347
18	30	15%	150	922	52	500,552 ± 4,025	496,398 ± 10,633	11,190	29,563	10,503 ± 28,874	14,657
18	30	15%	200	1,353	53	498,422 ± 2,927	496,136 ± 10,855	8,138	30,180	11,496 ± 27,149	13,782
20	30	5%	50	300	8.1	405,455 ± 642	405,190 ± 3,561	1,784	9,900	3,937 ± 8,278	4,202
20	30	5%	100	978	18	405,002 ± 679	402,537 ± 2,430	1,887	6,755	643 ± 6,123	3,108
20	30	5%	150	1,130	27	406,240 ± 719	402,497 ± 2,962	1,998	8,236	-62 ± 7,251	3,681
20	30	5%	200	1,593	42	406,393 ± 759	402,563 ± 2,494	2,111	6,933	-577 ± 6,408	3,253
20	30	10%	50	245	7.4	417,485 ± 1,573	417,721 ± 8,295	4,373	23,063	10,104 ± 19,439	9,868
20	30	10%	100	1,018	29	420,103 ± 2,326	416,029 ± 7,621	6,467	21,188	5,872 ± 19,594	9,946
20	30	10%	150	1,497	43	417,842 ± 642	416,540 ± 5,160	1,785	14,347	4,500 ± 11,430	5,802
20	30	10%	200	2,751	74	419,555 ± 2,063	417,450 ± 5,057	5,735	14,061	5,015 ± 14,026	7,120
20	30	15%	50	468	12	433,294 ± 1,794	434,664 ± 11,949	4,989	33,224	15,114 ± 27,075	13,744
20	30	15%	100	1,166	30	432,902 ± 1,455	432,580 ± 9,306	4,046	25,874	10,439 ± 21,199	10,761
20	30	15%	150	897	32	434,055 ± 1,025	432,326 ± 9,826	2,850	27,319	9,122 ± 21,375	10,851
20	30	15%	200	1,412	41	433,183 ± 669	432,376 ± 8,862	1,859	24,639	8,723 ± 18,774	9,530

Table C.1: Sample Average Approximation Method (continue)

$\lvert\mathcal{G}\rvert$	**M**	$\sigma_{\mathbf{D}}$	**N**	**Time (LB)**	**Time (UB)**	**LB**	**UB**	$\sigma_{\mathbf{LB}}$	$\sigma_{\mathbf{UB}}$	**GAP**	$\sigma_{\mathbf{GAP}}$
22	30	5%	50	215	4	267,285 ± 1,048	265,765 ± 2,099	2,915	5,835	1,627 ± 6,200	3,147
22	30	5%	100	780	16	266,507 ± 494	261,548 ± 2,105	1,372	5,854	-2,360 ± 5,120	2,599
22	30	5%	150	872	22	266,551 ± 245	266,642 ± 2,433	680	6,765	2,769 ± 5,275	2,678
22	30	5%	200	1,323	37	266,617 ± 313	265,931 ± 2,424	870	6,739	2,051 ± 5,391	2,737
22	30	10%	50	250	5	278,390 ± 939	276,862 ± 3,884	2,611	10,800	3,295 ± 9,502	4,823
22	30	10%	100	854	15	279,229 ± 750	277,883 ± 8,592	2,084	23,889	7,996 ± 18,402	9,342
22	30	10%	150	961	24	278,903 ± 919	271,473 ± 4,694	2,555	13,050	-1,817 ± 11,056	5,613
22	30	10%	200	1,422	37	280,282 ± 284	280,109 ± 3,303	789	9,183	3,414 ± 7,065	3,587
22	30	15%	50	274	5	298,450 ± 1,570	298,466 ± 14,412	4,364	40,070	15,997 ± 31,482	15,981
22	30	15%	100	989	19	299,453 ± 1,313	289,070 ± 8,728	3,652	24,267	-342 ± 19,781	10,041
22	30	15%	150	1,067	27	299,456 ± 1,060	294,385 ± 5,171	2,948	14,377	1,160 ± 12,275	6,231
22	30	15%	200	1,568	36	297,385 ± 1,586	284,839 ± 9,265	4,410	25,759	-1,695 ± 21,375	10,851
24	30	5%	50	277	7	267,606 ± 1,073	266,434 ± 2,898	2,983	8,057	2,799 ± 7,822	3,971
24	30	5%	100	1,170	19	266,728 ± 733	265,769 ± 2,195	2,037	6,103	1,969 ± 5,767	2,928
24	30	5%	150	1,302	30	266,462 ± 277	266,610 ± 2,432	769	6,763	2,857 ± 5,337	2,709
24	30	5%	200	2,223	51	266,483 ± 346	265,617 ± 2,501	963	6,953	1,981 ± 5,609	2,847
24	30	10%	50	441	8	278,591 ± 930	277,523 ± 4,241	2,587	11,791	4,103 ± 10,187	5,171
24	30	10%	100	1,582	15	279,482 ± 814	274,220 ± 6,499	2,263	18,071	2,051 ± 14,407	7,313
24	30	10%	150	1,172	26	279,128 ± 633	276,603 ± 3,985	1,761	11,080	2,093 ± 9,098	4,618
24	30	10%	200	2,291	51	280,008 ± 1,443	276,490 ± 3,913	4,011	10,880	1,838 ± 10,551	5,356
24	30	15%	50	451	9	299,097 ± 1,552	300,159 ± 13,627	4,316	37,888	16,241 ± 29,902	15,179
24	30	15%	100	1,539	21	298,118 ± 1,894	286,981 ± 11,804	5,267	32,820	2,561 ± 26,985	13,698
24	30	15%	150	1,401	31	299,583 ± 1,314	286,991 ± 11,687	3,653	32,495	409 ± 25,611	13,001
24	30	15%	200	1,937	41	298,623 ± 1,337	285,805 ± 10,900	3,718	30,306	-581 ± 24,107	12,237

Table C.1: Sample Average Approximation Method (continue)

List of Tables

Acknowledgements

I would like to express my deepest appreciation to my friends and family that constantly push me forward and provide me with their full support and encouragement.

This endeavour could not have been possible without my supervisor and co-supervisor patience and hard work. I'm extremly grateful to them.

And for to my life-mate, I am deeply in debt for holding me in the most challenging times.

I am truly grateful to you all.