DATA-DRIVEN RISK-AVERSE STOCHASTIC PROGRAM AND RENEWABLE ENERGY INTEGRATION

By

CHAOYUE ZHAO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2014

To my parents Yongchun Zhao and Yajing Ren

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor Dr. Yongpei Guan, for his constant support, guidance, encouragement and friendship for the past four years. I am more than grateful to have the invaluable opportunity to work with him, and to learn from him. His knowledge, passion, wisdom and enthusiasm help mold me into the researcher I am today and inspire me to hopefully become a professor like him in future. Without his guidance and supervision, this dissertation work would never be possible.

I am very grateful to Dr. Joseph Geunes, Dr. Jean-Philippe Richard and Dr. William Hager for being on my PhD dissertation committee, for their valuable comments and suggestions on this dissertation, and for their sincere support, suggestions and guide on me. I also appreciate their responsiveness and flexibility to support my compressed timeline.

I am also grateful to the staff at Department of Industrial and Systems Engineering, Ms. Cynthia Blunt, Ms. Sara Pons, for their administrative supports.

In addition, I would like to thank my collaborators: Jianhui Wang and Jean-Paul Watson, for their advices and suggestions. It has been a great experience to work with them. Many thanks to my amazing graduate student colleagues and friends at University of Florida, it is your friendship that made my graduate study one of the best experiences of my life.

Finally, I deeply appreciate my parents Yongchun Zhao, Yajing Ren, and my boyfriend Yuanxiang Wang, for their encouragement, support and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

DATA-DRIVEN RISK-AVERSE STOCHASTIC PROGRAM AND RENEWABLE ENERGY INTEGRATION

By

Chaoyue Zhao

August 2014

Chair: Yongpei Guan
Major: Industrial and Systems Engineering

With increasing penetration of renewable energy into the power grid and its intermittent nature, it is crucial and challenging for system operators to provide reliable and cost effective daily electricity generation scheduling. In this dissertation, we present our recently developed innovative modeling and solution approaches to address this challenging problem. We start with developing several optimization-under-uncertainty models, including both stochastic and robust optimization ones, to solve reliability unit commitment problems for Independent System Operators (ISOs) so as to ensure power system cost efficiency while maintaining a high utilization of renewable energy. Then, we extend our research to study data-driven risk-averse two-stage stochastic program, for which the distribution of the random variable is within a given confidence set. By introducing a new class of probability metrics, we construct the confidence set based on historical data, and provide a framework to solve the problem for both discrete and continuous distribution cases. Our approach is guaranteed to obtain an optimal solution and in addition, we prove that our risk-averse stochastic program converges to the risk-neutral case as the size of historical data increases to infinity. Moreover, we show the "value of data" by analyzing the convergence rate of our solution approach. Finally, we illustrate examples of using this framework and discuss its application on renewable energy integration.

CHAPTER 1
INTRODUCTION

The Unit Commitment (UC) problem is one of the most important problems in power systems, which evolves determining effective on/off daily scheduling of generation units, to satisfy the forecasted electricity consumption over a given time horizon, while adhering to the units' physical and transmission constraints at minimum operation cost. The "unit commitment" decision is to determine the start-up and shut-down schedule of generation units of each hour over the planning time horizon while satisfying the start-up/shut-down and on/off status constraints. The "economic dispatch" decision is the one that can make the electricity supply meet the electricity load simultaneously, with consideration of generation and transmission constraints.

In the wholesale electricity markets, an Independent System Operator (ISO) accepts bids from both supply and load sides, and runs the unit commitment problem to obtain the generation schedule, to keep the power system balanced. However, the increasing integration of intermittent renewable energy into the power system brings a high degree of uncertainty to both the supply side and the load side, and consequently brings profound challenges to the ISO in order to maintain a stable and reliable power grid. On one hand, government policies encourage the large scale penetration of renewable energy, which leads to positive impacts on greenhouse gas reduction, water conservation, and energy security. For instance, the system operators in some regions, such as Germany, consider renewable energy as a higher priority over other conventional generation sources [23]. Besides, the California Public Utilities Commission enforces at least 33 percentages of the electricity retail sales from the renewable energy by 2020 [39]. On the other hand, due to the intermittent nature of renewable energy, it is very difficult to accurately predict its output. The inaccurate forecast of renewable energy output introduces considerable uncertainties into the system operator's decision-making process. If the uncertainties cannot be

handled well, transmission violations, load blackouts and even cascading failures may occur in real time, which could cause high-priced remedy actions to correct potential system imbalance. Over the years, industry puts significant efforts on dealing with uncertainties in the power system and has developed many effective ways to handle the uncertainties. One way to prevent the system imbalance caused by the uncertain system environment is to reserve generation capacity, e.g., the operating reserves. For example, at MISO, the operating reserves include regulating reserve, spinning reserve, and supplemental reserve [15]. Although operating reserves are regarded as effective ways to accommodate uncertainties, they are not sufficient to cover larger uncertainties, such as renewable energy output uncertainties. Recently, two-stage stochastic and robust optimization approaches have been studied extensively to accommodate uncertainties.

For the two-stage stochastic optimization approach, the day-ahead unit commitment decision is made in the first stage before the uncertain problem parameter representing the real time information is realized and the economic dispatch amount is made in the second stage after the uncertain parameter is realized. The objective is to minimize the total expected cost and the uncertain problem parameter (e.g., wind power) is captured by a number of scenarios. In recent works, significant contribution has been made by using the stochastic optimization models to solve the unit commitment problem under uncertainties, in particular, under wind power output uncertainty. For instance, recently in [5], [64], and [63], a stochastic unit commitment model is introduced for short-term operations to integrate wind power in the Liberalised Electricity Markets. This model has been successfully implemented and used in several wind power integration studies. In addition, a stochastic formulation, which allows the explicit modeling of the sources of uncertainty in the unit commitment problem, is proposed in [55], a two-stage security-constrained unit commitment (SCUC) algorithm that considers the unit commitment decision in the first stage and takes into account the

intermittency and volatility of wind power generation in the second stage is introduced in [70], and a stochastic unit commitment model, considering various wind power forecasts and their impacts on unit commitment and dispatch decisions, is proposed in [69]. Significant research progress has also been made to solve security-constrained stochastic unit commitment models. For instance, security-constrained stochastic unit commitment formulations addressing market-clearing are described in [13] and the corresponding case studies are performed in [14]. In [75], a scenario-tree based stochastic security-constrained unit commitment is studied. Most recently, two-stage stochastic programming approaches have been studied to consider both slow-start and fast-start generators in [49], in which the slow-start generators are committed in the first stage and fast-start generators are committed in the second stage. These approaches have also been studied to estimate the contribution of demand flexibility in replacing operating reserves in [48] and ensure high utilization of wind power output by adding additional chance constraints in [71].

In practice, a significant amount of available data for ISOs/RTOs makes it possible to take samples and generate scenarios for the stochastic optimization approach. However, it is always challenging for the stochastic optimization approach to deal with large-sized instances when the scenario size increases significantly. Therefore, different scenario reduction approaches have been proposed to select important scenarios. In this way, the small sample size may lead to the feasibility issues. That is, the day-ahead unit commitment decision might not be feasible for some scenarios which are not selected. Recently, robust optimization approaches have been proposed to ensure the robustness and make the day-ahead unit commitment feasible for most outcomes of the real time uncertain input parameter. For the robust optimization approach, the uncertain parameter is described within a given deterministic uncertainty set and the objective is to minimize the worst-case cost that includes the first-stage unit commitment and the second-stage economic dispatch costs. Recent research works include two-stage

robust unit commitment model and Benders' decomposition algorithm developments to ensure system robustness under load uncertainties introduced in [32] and [7], two-stage robust optimization models with slightly different uncertainty sets to provide robust day-ahead unit commitment decisions under wind power output uncertainties presented in [78] and [30], a robust bidding strategy in a pool-based market by solving a robust mixed-integer linear program and generating a bidding curve described in [5], a robust optimization model to integrate PHEVs into the electric grid to handle the most relevant planning uncertainties proposed in [28], and a robust optimization approach to solve contingency-constrained unit commitment with N-k security criterion and decide reserve amounts to ensure system robustness introduced in [59].

The advantage of the robust optimization approach is that it requires minimal information of the input uncertain parameter (as long as the information is sufficient to generate the deterministic uncertainty set) and ensures the robustness of the obtained unit commitment decision, i.e., the day-ahead unit commitment decision is feasible for most outcomes of the real time uncertain problem parameter. However, this approach always faces the challenges on its over conservatism, due to its objective function of minimizing the worst-case cost, because the worst case happens rarely.

In this dissertation, we present our recently developed innovative modeling and solution approaches to address the unit commitment problem. We start with developing several optimization-under-uncertainty models, including both stochastic and robust optimization ones, to solve reliability unit commitment problems for ISOs so as to ensure power system cost efficiency while maintaining a high utilization of renewable energy. In addition, to address the shortages of stochastic optimization and robust optimization approaches, we propose an innovative unified stochastic and robust unit commitment model to take advantage of both stochastic and robust optimization approaches. Then, we extend our research to study the data-driven risk-averse two-stage stochastic program, in which the distribution of the random variable is within a given confidence

set. Finally, we apply the proposed data-driven risk-averse two-stage stochastic optimization framework to power system problems. In this dissertation, we unfold the discussion as follows.

In Chapter 2, we describe a multi-stage robust unit commitment problem considering wind and demand response (DR) uncertainties. In this chapter, DR programs are introduced as reserve resource to mitigate wind power output uncertainty. However, the price-elastic demand curve is not exactly known in advance, which provides another dimension of uncertainty. To accommodate the combined uncertainties from wind power and DR, we allow the wind power output to vary within a given interval with the price-elastic demand curve also varying in this dissertation. We develop a robust optimization approach to derive an optimal unit commitment decision for the reliability unit commitment runs by ISOs/RTOs, with the objective of maximizing total social welfare under the joint worst-case wind power output and demand response scenario. The problem is formulated as a multi-stage robust mixed-integer programming problem. An exact solution approach leveraging Benders' decomposition is developed to obtain the optimal robust unit commitment schedule for the problem. Additional variables are introduced to parameterize the conservatism of our model and avoid over-protection. Finally, we test the performance of the proposed approach using a case study based on the IEEE 118-bus system. The results verify that our proposed approach can accommodate both wind power and demand response uncertainties, and demand response can help accommodate wind power output uncertainty by lowering the unit load cost.

In Chapter 3, we propose a novel unified stochastic and robust unit commitment model that takes advantage of both stochastic and robust optimization approaches, that is, this innovative model can achieve a low expected total cost while ensuring the system robustness. By introducing weights for the components for the stochastic and robust parts in the objective function, system operators can adjust the weights based

on their preferences. Finally, a Benders' decomposition algorithm is developed to solve the model efficiently. The computational results indicate that this approach provides a more robust and computationally trackable framework as compared to the stochastic optimization approach and a more cost-effective unit commitment decision as compared to the robust optimization approach.

Chapter 4 presents a data-driven risk-averse two-stage stochastic optimization framework. In most practice, the actual distribution of a random parameter is unknown. Instead, only a series of historic data are available. In this chapter, we develop a data-driven stochastic optimization approach to provide a risk-averse decision making under uncertainty. In our approach, starting from a given set of historical data, we first construct a confidence set for the unknown probability distribution utilizing a class of ζ-structure probability metrics. Then, we describe the reference distributions and solution approaches to solve the developed two-stage risk-averse stochastic program, corresponding to the given set of historical data, for the cases in which the true probability distributions are discrete and continuous, respectively. More specifically, for the case in which the true probability distribution is discrete, we reformulate the risk-averse problem to a traditional two-stage robust optimization problem. For the case in which the true probability distribution is continuous, we develop a sampling approach to obtain the upper and lower bounds for the risk-averse problem, and prove that these two bounds converge to the optimal objective value exponentially fast as the sample size increases. Furthermore, we prove that, for both cases, as more data samples are observed, the risk-averse problem converges to the risk-neutral one exponentially fast as well. Finally, the experimental results on newsvendor and facility location problems show how numerically the optimal objective value of the risk-averse stochastic program converges to the risk-neutral one, which indicates the value of data.

Chapter 5 utilizes the data-driven risk-averse two-stage stochastic optimization techniques to solve the unit commitment problem under renewable energy output

uncertainty. The traditional way to describe the renewable energy output uncertainty is to assume the output follows a certain distribution, e.g., Normal distribution and Weibull distribution. However, in practice, the information about the renewable energy output is usually incomplete, and the inaccurate distribution assumption may lead to biased UC and ED solutions. In this chapter, instead of assuming the distribution of renewable energy output is known, a series of historical data, which are drawn from the true distribution, are observed. We construct the confidence set of the renewable energy output distribution, and propose a risk-averse approach to address the unit commitment problem under renewable energy output uncertainty. That is, we consider the worst case distribution in the confidence set. We introduce the Wasserstein distribution metric and propose the solution approaches to tackle the data-driven risk-averse stochastic optimization framework. We prove that as the number of historical data goes to infinity, the risk-averse solution converges to the risk-neutral solution. Our case study on the IEEE 118-bus system verifies the effectiveness of our proposed solution approach.

Finally, Chapter 6 concludes the dissertation and provides general suggestions for future research.

CHAPTER 2
MULTI-STAGE ROBUST UNIT COMMITMENT CONSIDERING WIND AND DEMAND RESPONSE UNCERTAINTIES

2.1 Problem Description and Literature Review

In recent years, wind energy penetration has increased substantially and is expected to continue growing in the future. For example, the U.S. Department of Energy described a scenario that wind energy could generate 20% of nation's electricity by 2030 [44]. However, due to its intermittent nature, wind power is inherently very difficult to predict. Moreover, the magnitude of wind power output variance is much larger than that of the traditional load variance. As a result, traditional power system operation methods are insufficient to maintain system reliability. Due to physical constraints of the power system (e.g., ramping limits of conventional generators and transmission line capacities), wind power curtailment occurs frequently, which consequently leads to low utilization of wind power and dampens the incentive of wind power investment in the long run. Therefore, the system operators in some regions, such as Germany, consider renewable energy a higher priority over the other conventional generation sources [23].

Recent studies have focused on developing stochastic optimization models with the objective of minimizing the total expected cost, including a short-term stochastic rolling unit commitment model [5, 63], a stochastic unit commitment model to calculate reserve requirements by simulating the wind power realizations and comparing its performance with the traditional pre-defined reserve requirements [12], and a study on the impacts of wind power on thermal generation unit commitment [65]. Related research can also be found in [55], [70], and [69]. All of this research indicates that wind power output uncertainty and wind power forecast errors have a significant impact on unit commitment

and dispatch, and more advanced power system operation methods are required to make the system reliable.

More recently, to ensure high utilization of wind power, a chance constrained optimization model [71] and a robust optimization model [31] have been developed to solve the problem. In the former case, a chance constraint is developed to ensure that a portion of the wind power output (e.g., 90%) be utilized at a certain probability. In this way, the risk of a large amount of wind power being curtailed can be adjusted by the system operators. In the latter case, wind power output is assumed to lie within an interval defined by quantiles. All of the wind power output within this interval will be utilized. Unit commitment decisions are then made by considering the worst-case wind power output scenario. While these two approaches ensure high utilization of wind power output, both approaches tend to commit more conventional generators to accommodate the wind power output uncertainty. As an alternative, a pumped storage hydro unit is considered in [31]. Pumped storage hydro is flexible and easy to operate, and can reduce the total cost under the worst-case scenario significantly. However, due to locational and geographical limitations, pumped storage hydro units can not be widely adopted.

In comparison, demand response (DR) has been shown to be an efficient approach to reduce peak load [60] [35]. It also has potential to accommodate wind power output uncertainty. For instance, when the wind power output is higher than expected, DR programs can help absorb the extra wind power. On the other hand, DR programs can help decrease the load when the wind power output is low. More importantly, this approach can be widely applied, as compared to pumped storage hydro.

In general, DR aims to manage end-use consumers' electricity consumption patterns via time-varying prices, or by offering financial incentives to reduce the consumption of electricity at times of high electricity prices or when system reliability is jeopardized [43]. DR can benefit load-serving entities, consumers, and Independent

System Operators (ISOs) [60] [35] [43]. In particular, for ISOs (the focus of this chapter), DR can help balance electricity consumption and generation, and can therefore ensure a more stable, reliable and controllable power grid.

The U.S. Department of Energy predicted that by 2019, the total U.S. peak demand could be reduced 20% by DR with full participation [41]. In order to "ensure that demand response is treated comparably to other resources," the Federal Energy Regulatory Commission (FERC) requires that ISOs and Regional Transmission Organizations (RTOs) "accept bids from demand response resources in their markets for certain ancillary services, comparable to other resources" [40]. Several regional grid operators (e.g., NYISO, PJM, ISO-NE, and ERCOT) have provided opportunities for consumers to participate in DR programs in order to integrate DR resources into the wholesale energy market step by step.

In most research, the price-elastic demand curve is characterized by price elasticity, which represents the sensitivity of electricity demand (Q) with respect to the change of price (P) [2, 61]. For a small change in price (ΔP), price elasticity is defined as

$$\alpha = \frac{\Delta Q/Q}{\Delta P/P}. \tag{2–1}$$

In [60], the elasticity value is simplified as $\alpha = \Delta Q/\Delta P$, resulting in a linearized price-elastic demand curve. In [34], the price-elastic demand curve is approximated as a stepwise linear curve. In [36], provided that the change in price of one commodity will not only affect its demand, but also may affect the demand of another commodity, the concept of "self-elasticity" and "cross-elasticity" is developed. The chapter also analyzes how these elasticities can model consumers' behaviors and the set of spot prices.

In the above research, DR was mostly modeled as a fixed price-elastic demand curve. However, due to a variety of reasons including lack of attention, latency in communication, and change in consumption behavior, the actual price-elastic demand curve is uncertain in nature [47]. In other words, the actual response from

the consumers in real time could be different from the forecasted values. Therefore, the consumer behavior should be modeled by an uncertain price-elastic demand curve, which means consumers have different response patterns to the electricity prices under different scenarios. In this case, the price-elastic demand curve can vary within a certain range. Hence, we propose an efficient robust unit commitment approach that can consider wind power output uncertainty and inexact DR information in this chapter. We assume wind power output is within a given interval and the price-elastic demand curve is also varying within a given range. The objective is to maximize the social welfare (defined in Section 2.3) under the worst-case joint wind power output and price-elastic demand curve scenario. Our first stage variables are unit commitment decisions; the second stage considers economic dispatch for each thermal generator after the worst-case wind power output scenario is realized; in the third stage, we consider the worst-case price-elastic demand curve. By using this robust optimization approach, the reliability unit commitment run process (e.g., referred as reliability unit commitment at ERCOT and reliability assessment commitment at Midwest ISO) at each ISO can be strengthened. As compared to the recent works on robust optimization to solve power system optimization problems [4, 28, 31, 59], the contributions of this chapter can be summarized as follows:

1. Both wind power output and demand response uncertainties are considered in the unit commitment problem.
2. A multi-stage robust optimization model is developed to formulate the problem, as compared to previously studied two-stage robust optimization models.
3. A tractable solution approach is proposed to solve the multi-stage robust optimization problem and the computational results verify the effectiveness of our proposed approach.

The remainder of the chapter is organized as follows. In Section 2.3, we describe how to formulate the uncertainty sets describing the uncertain wind power output and the uncertain region of the price-elastic demand curve. We then derive a multi-stage

robust optimization model to decide the optimal robust unit commitment schedule. In Section 2.4, we take advantage of the problem structure and transform the multi-stage robust optimization problem into a two-stage problem. Then, we develop a Benders' decomposition algorithm to solve the problem. In Section 2.5, we provide case studies and examine associated computational results. We conclude with a summary of our contributions and discussions in Section 2.6.

2.2 Nomenclature

A. Sets and Parameters

T Set of time periods.

B Set of buses.

K Set of steps introduced to approximate the price-elastic demand curve.

G_b Set of generators at bus b.

Ω Set of transmission lines linking two buses.

SU_i^b Start-up cost for generator i at bus b.

SD_i^b Shut-down cost for generator i at bus b.

MU_i^b Minimum up-time for generator i at bus b.

MD_i^b Minimum down-time for generator i at bus b.

L_i^b Lower bound of electricity generated by thermal generator i at bus b.

U_i^b Upper bound of electricity generated by thermal generator i at bus b.

RU_i^b Ramp-up rate limit for generator i at bus b.

RD_i^b Ramp-down rate limit for generator i at bus b.

C_{ij} Transmission capacity for the transimission line linking bus i and bus j.

K_{ij}^b Line flow distribution factor for the transimission line linking bus i and bus j, due to the net injection at bus b, as described in [73].

D_{tb}^0 The inelastic part of demand at bus b in time period t.

D_{tb}^{M} The maximum demand at bus b in time period t.

W_t^{b*} The forecasted wind power output at bus b in time period t.

W_t^{b+} The upper deviation of the confidence interval for the wind power output at bus b in time period t.

W_t^{b-} The lower deviation of the confidence interval for the wind power output at bus b in time period t.

ϖ_t^b The upper bound for the total deviations of the real price-elastic demand curve from the forecasted curve at bus b in time period t.

π^b The cardinality budget to restrict the number of time periods in which the wind power output is far away from its forecasted value at bus b.

ℓ_t^{bk} The kth step length in the price-elastic demand curve at bus b in time period t.

p_t^{bk} The price at step k in the price-elastic demand curve at bus b in time period t.

α_t^b The given price elasticity at bus b in time period t.

B. Decision Variables

y_{it}^b Binary variable to indicate if generator i is on at bus b in time period t.

u_{it}^b Binary variable to indicate if generator i is started up at bus b in time period t.

v_{it}^b Binary variable to indicate if generator i is shut down at bus b in time period t.

w_t^b Wind power output at bus b in time period t.

d_t^b Actual electricity demand at bus b in time period t.

x_{it}^b Amount of electricity produced by generator i at bus b in time period t.

h_t^{bk} The auxillary variable introduced for demand at step k in the price-elastic demand curve at bus b in time period t.

$r_t^b(.)$ The integral of the price-elastic demand curve at bus b in time period t.

$f_i^b(.)$ The fuel cost function of generator i at bus b.

C. Random Parameters

p_t^b Electricity price at bus b in time period t.

2.3 Mathematical Formulation

In this section, we first describe the deterministic model for ISOs/RTOs to determine unit commitment decisions with the objective of maximizing total social welfare. In this model, the wind power output w_t^b is assumed deterministic and the price-elastic demand curve is also certain. In general, demand will decrease when electricity price increases. However, some electricity consumption will not be affected by electricity prices, such as critical loads like hospitals and airports. We define this part of demand as "inelastic demand". Accordingly, the other part of demand varying with electricity prices is referred to as "elastic demand" [60].

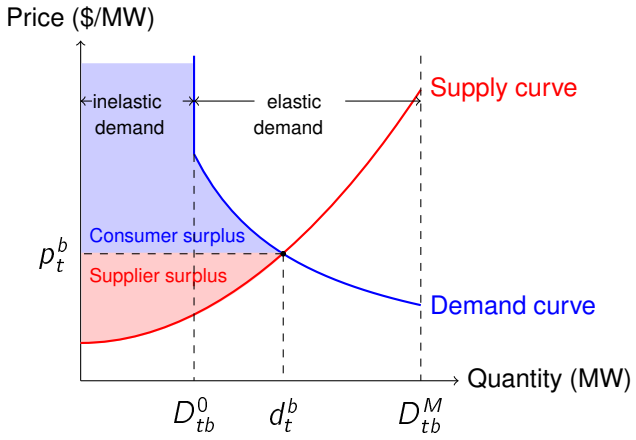

Figure 2-1. An example of price-elastic demand curve

In this chapter, we assume load at each bus includes both inelastic and elastic components. We can model the demand curve and supply curve as shown in Fig. 2-1. The electricity supply and demand reach an equilibrium at the intersection point (d_t^b, p_t^b), corresponding to the electricity demand d_t^b leading to the maximum social welfare, which is defined as the summation of consumer surplus and supplier surplus as shown in Fig. 2-1. Since the inelastic demand part has an infinite marginal value, we assume that the consumer surplus for the inelastic demand part is a constant as defined in [60]. In addition, for our price-elastic demand curve, we assume the elastic load is a non-increasing function of the electricity price (cf. [60], [36], and [42]). Let D_{tb}^0 be

the inelastic demand at bus b in time period t. Because the demand has the inelastic component, we have $d_t^b \geq D_{tb}^0$. We further impose an upper limit D_{tb}^M on d_t^b, yielding: $D_{tb}^0 \leq d_t^b \leq D_{tb}^M$. Accordingly, the social welfare is equal to the difference between the integral of the demand curve from D_{tb}^0 to d_t^b (denoted as $r_t^b(d_t^b)$ in our model) plus a constant (i.e., the integral of the demand curve from 0 to D_{tb}^0) and the integral of the supply curve from 0 to d_t^b. In our model, for the calculation convenience, we omit the constant part in our objective function, which will provide the same optimal solution. Finally, we let $f_i^b(x_{it}^b)$ represent the fuel cost for generator i at bus b in time period t. The deterministic model can be described as follows (denoted as D-UC):

2.3.1 Deterministic Model

$$\max \sum_{t\in T}\sum_{b\in B} r_t^b(d_t^b) - \sum_{t\in T}\sum_{b\in B}\sum_{i\in G_b}(f_i^b(x_{it}^b) + SU_i^b u_{it}^b + SD_i^b v_{it}^b) \tag{2–2}$$

$s.t.$

$$-y_{i(t-1)}^b + y_{it}^b - y_{ik}^b \leq 0, \tag{2–3}$$
$$\forall k : 1 \leq k - (t-1) \leq MU_i^b, \forall t \in T, \forall b \in B, \forall i \in G_b$$

$$y_{i(t-1)}^b - y_{it}^b + y_{ik}^b \leq 1, \tag{2–4}$$
$$\forall k : 1 \leq k - (t-1) \leq MD_i^b, \forall t \in T, \forall b \in B, \forall i \in G_b$$

$$-y_{i(t-1)}^b + y_{it}^b - u_{it}^b \leq 0, \quad \forall t \in T, \forall b \in B, \forall i \in G_b \tag{2–5}$$

$$y_{i(t-1)}^b - y_{it}^b - v_{it}^b \leq 0, \quad \forall t \in T, \forall b \in B, \forall i \in G_b \tag{2–6}$$

$$L_i^b y_{it}^b \leq x_{it}^b \leq U_i^b y_{it}^b, \quad \forall t \in T, \forall b \in B, \forall i \in G_b \tag{2–7}$$

$$x_{it}^b - x_{i(t-1)}^b \leq y_{i(t-1)}^b RU_i^b + (1 - y_{i(t-1)}^b) MA_i^b, \tag{2–8}$$
$$\forall t \in T, \forall b \in B, \forall i \in G_b$$

$$x_{i(t-1)}^b - x_{it}^b \leq y_{it}^b RD_i^b + (1 - y_{it}^b) MA_i^b, \tag{2–9}$$
$$\forall t \in T, \forall b \in B, \forall i \in G_b$$

$$\sum_{b\in B}(\sum_{i\in G_b} x_{it}^b + w_t^b) = \sum_{b\in B} d_t^b, \quad \forall t \in T \tag{2–10}$$

$$-C_{ij} \le \sum_{b\in B} K_{ij}^b (\sum_{q\in G_b} x_{qt}^b + w_t^b - d_t^b) \le C_{ij}, \tag{2–11}$$
$$\forall (i,j) \in \Omega, \forall t \in T$$

$$D_{tb}^0 \le d_t^b \le D_{tb}^M, \quad \forall t \in T, \forall b \in B \tag{2–12}$$

$$y_{it}^b, u_{it}^b, v_{it}^b \in \{0,1\}, x_{it}^b, d_t^b \ge 0, \forall t \in T, \forall b \in B, \forall i \in G_b. \tag{2–13}$$

In the above formulation, the objective function in (2–2) is to maximize the social welfare (without the constant part). Constraints (2–3) and (2–4) represent the minimum up-time and the minimum down-time restrictions respectively. Constraints (2–5) and (2–6) compute the unit start-up and shut-down state variables. Constraints (2–7) enforce the upper and lower limits of power output of each unit. Constraints (2–8) and (2–9) enforce the ramping rate limits of each unit. Constraints (2–10) ensure load balance and require the supply to meet the demand. Constraints (2–11) are transmission line capacity limits. Finally, constraints (2–12) enforce the lower and upper limits for actual demand, due to the contribution of elastic demand [61].

In the following subsections, we describe how to linearize the objective function, generate the uncertainty sets for the wind power output and price-elastic demand curve, and obtain a final robust optimization formulation.

2.3.2 Linearizing the Objective Function

There are two nonlinear terms in the objective function for (D-UC): $r_t^b(d_t^b)$ and $f_i^b(x_{it}^b)$. Now we describe how to linearize these two nonlinear terms.

2.3.2.1 Linearizing $r_t^b(d_t^b)$

The elasticity value determines the flexibility of the demand. We model the consumer demand response of the "elastic demand" part as a price-elastic demand curve. If price elasticity is constant, we can represent the price-elastic demand curve as:

$$d_t^b = A_t^b (p_t^b)^{\alpha_t^b}, \tag{2–14}$$

where α_t^b is the given price elasticity at bus b in time period t, and A_t^b is a parameter that can be decided by a given reference point (d_t^{b*}, p_t^{b*}) [61].

Finally, note here that our proposed solution method can be applied to other modeling approaches of demand elasticity without loss of generality. For instance, for some ISOs/RTOs (e.g.,MISO), the price-elastic demand curve itself is a step-wise function. As shown in Fig. 2-2, for a general price-elastic demand curve represented as (2–14), a step-wise function can be applied to approximate this price-elastic demand curve. We can approximate $r_t^b(d_t^b)$ as:

$$r_t^b(d_t^b) = \sum_{k \in K} p_t^{bk} h_t^{bk}, \tag{2–15}$$

$$d_t^b = \sum_{k \in K} h_t^{bk}, \tag{2–16}$$

$$0 \leq h_t^{bk} \leq l_t^{bk}, \forall k \in K \tag{2–17}$$

where l_t^{bk} is the kth step length of the step-wise function, p_t^{bk} is the corresponding price at step k, h_t^k is the auxiliary variable introduced for demand at step k, and K is the set of steps.

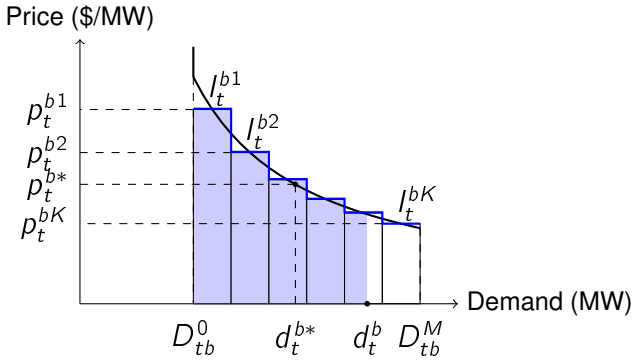

Figure 2-2. Step-wise function approximation of the price-elastic demand curve

Notice that p_t^{bk} is strictly decreasing with k. Since we are maximizing $r_t^b(d_t^b)$, according to equations (2–16) and (2–17), we have:

$$h_t^{bs} = \begin{cases} l_t^{bs}, & \text{if } s < s_0; \\ d_t^b - \sum_{k=1}^{s_0-1} l_t^{bk}, & \text{if } s = s_0; \\ 0, & \text{if } s > s_0, \end{cases}$$

when $\sum_{k=1}^{s_0-1} l_t^{bk} < d_t^b \leq \sum_{k=1}^{s_0} l_t^{bk}$ for a certain s_0. In this case, we can prove that $\sum_{k\in K} p_t^{bk} h_t^{bk}$ is the approximated integral of the price-elastic demand curve, i.e., equation (2–15) is justified.

2.3.2.2 Linearizing $f_i^b(x_{it}^b)$

In practice, the fuel cost function $f_i^b(x_{it}^b)$ can be expressed as a quadratic function, which we will use the following N-piece piecewise linear function to approximate [31]:

$$\phi_{it}^b \geq \gamma_{it}^{jb} y_{it}^b + \beta_{it}^{jb} x_{it}^b, \tag{2–18}$$
$$\forall t \in T, \forall b \in B, \forall i \in G_b, \forall j = 1, \cdots, N$$

where γ_{it}^{jb} and β_{it}^{jb} are the intercept and the slope of the jth segment line and ϕ_{it}^b is an auxiliary variable.

2.3.3 Uncertain Wind Power Output Formulation

Due to the intermittent nature of wind, it is difficult to precisely characterize the wind power output. As shown in [31], we assume the wind power output is within an interval $[W_t^{b*} - W_t^{b-}, W_t^{b*} + W_t^{b+}]$ with W_t^{b*} representing the predicted value of the wind power output at bus b in time period t, and W_t^{b+}, W_t^{b-} representing the allowed maximum deviations above and below W_t^{b*}, respectively. This interval can usually be generated by using quantiles. For instance, we can set $W_t^{b*} + W_t^{b+}$ and $W_t^{b*} - W_t^{b-}$ equal to the $.95$- and $.05$-quantiles of the uncertain wind power output, respectively. The actual wind power output w_t^b is allowed to be any value in the given range. We use the cardinality uncertainty set to adjust the conservatism of our proposed model as shown in [8]. For this approach, we introduce an integer π^b as the *cardinality budget* [8] to restrict the number of time periods in which the wind power output is far away from its

forecasted value at bus b. The budget π^b can be adjusted by the system operators. For example, if π^b is set to be 0, the wind output fluctuation at each bus is assumed to be small and can be approximated by its forecasted value. If $\pi^b = 6$, significant fluctuations of wind power output are assumed to occur in no more than six time intervals. It can be observed that this "budget parameter" π^b can be used to adjust the conservatism of the system. For instance, our proposed approach is more conservatism as π^b increases. Meanwhile, it is proved in [8] that, for any given budget π^b less than 24, the robust optimal unit commitment solution obtained based on this uncertainty set is still feasible for any possible wind power output between its given lower and upper bounds with a high probability (e.g., when $\pi \geq 8$, the robust optimal unit commitment solution will be feasible with a probability higher than 95% as shown in [31]). Under this setting, at each bus b, the worst-case wind power output scenario happens when wind power output reaches its lower bound, upper bound or forecasted value, and the total number of periods in which wind power output is not at its forecasted value should be no more than the budget π^b. Accordingly, the uncertainty set can be described as follows:

$$\mathcal{W} := \Bigg\{ w \in \mathcal{R}^{|B|\times|T|} : w_t^b = W_t^{b*} + z_t^{b+} W_t^{b+} - z_t^{b-} W_t^{b-}, \\ \sum_{t\in T} (z_t^{b+} + z_t^{b-}) \leq \pi^b, \forall t \in T, \forall b \in B \Bigg\},$$

where z_t^{b+} and z_t^{b-} are binary variables. If $z_t^{b+} = 1$, the wind power output will reach its upper limit, and if $z_t^{b-} = 1$, the wind power output will reach its lower bound, and if both of them are 0, the forecasted value is achieved.

2.3.4 Uncertain Demand Response Curve Formulation

In Subsection 2.3.2.1, we provide a description of the price-elastic demand curve and how to use linear functions to approximate it. In practice, however, as described in Section 2.1, the actual price-elastic demand curve is uncertain. When ISOs/RTOs make unit commitment decisions, it is necessary to allow the price-elastic demand curve

to vary within a certain range. As shown in Fig. 2-3, for a given certain price p_0, the corresponding demand is uncertain (e.g., the range for d_0 as indicated in the figure). Similarly, for a given demand d_0, the corresponding price can vary within a certain range (e.g., the range for p_0 as indicated in the figure). Accordingly, we can formulate the

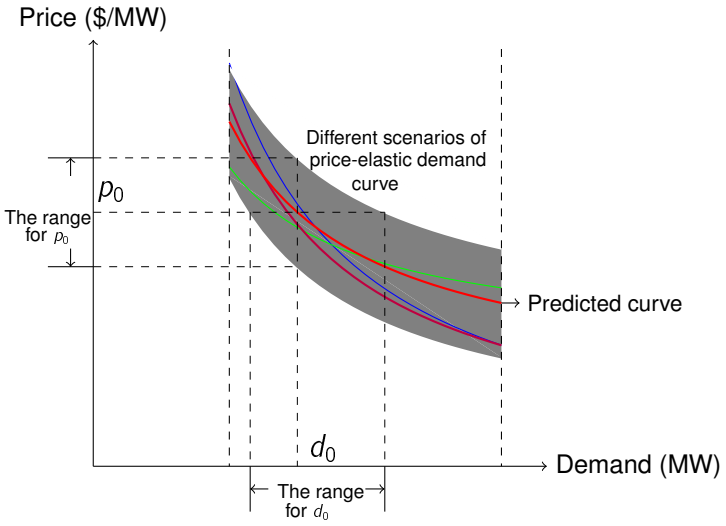

Figure 2-3. The uncertainty of price-elastic demand curve

price-elastic demand curve as $d_t^b = A_t^b (p_t^b)^{\alpha_t^b} + \epsilon_t^b$ or $d_t^b = A_t^b (p_t^b + \epsilon_t^b)^{\alpha_t^b}$, where ϵ_t^b represents a deviation used to describe the uncertainty of the price-elastic demand curve. In our model, we assume $d_t^b = A_t^b (p_t^b + \epsilon_t^b)^{\alpha_t^b}$ for computational convenience. Following the previous steps on approximating the price-elastic demand curve as a step-wise curve, for each d_t^b in the price-elastic demand curve, the corresponding p_t^{bk} is allowed to vary within the range $p_t^{bk} \in [p_t^{bk*} - \hat{\epsilon}_t^{bk}, p_t^{bk*} + \hat{\epsilon}_t^{bk}]$, where p_t^{bk*} represents the estimated value of p_t^{bk}, ϵ_t^{bk} is the deviation of p_t^{bk} and $\hat{\epsilon}_t^{bk}$ is the upper limit of ϵ_t^{bk}. To adjust the conservativeness, we introduce the parameter ϖ_t^b to restrict the total amount of deviations, i.e., $\sum_{k \in K} \epsilon_t^{bk}$. We can adjust the conservativeness of our proposed approach through changing the value of ϖ_t^b. The smaller the value ϖ_t^b is, the less uncertainty the demand response curve has. The uncertainty set of the demand

response curve can be described as follows:

$$\Pi = \Big\{\epsilon : \quad -\hat{\epsilon}_t^{bk} \leq \epsilon_t^{bk} \leq \hat{\epsilon}_t^{bk}, \tag{2–19}$$

$$-\varpi_t^b \leq \sum_{k\in K} \epsilon_t^{bk} \leq \varpi_t^b, \tag{2–20}$$

$$\forall t \in T, \forall b \in B, \forall k \in K\Big\}. \tag{2–21}$$

2.3.5 Robust Optimization Formulation

With the consideration of both wind and demand response uncertainties, we develop a three-stage robust optimization formulation to determine robust day-ahead reliability unit commitment decisions for ISOs/RTOs. In the first stage, we include the unit commitment decisions for each generation unit while considering all unit commitment constraints with unknown wind power output and demand response patterns. After realizing the worst-case wind power output, we decide in the second stage the dispatch level for each unit to maximize the total social welfare with the consideration of the worst-case price-elastic demand curve. Finally, in the third stage, we consider the uncertainty of the price-elastic demand curve that minimizes the expected total social welfare. The derived robust optimization formulation is shown as follows:

$$\begin{aligned}
&\max_{y,u,v}\Big\{ -\sum_{t\in T}\sum_{b\in B}\sum_{i\in G_b}(SU_i^b u_{it}^b + SD_i^b v_{it}^b) + \\
&\min_{w\in\mathcal{W}} \max_{x,h\in\chi}(\min_{\epsilon\in\Pi} \sum_{t\in T}\sum_{b\in B}\sum_{k\in K}(p_t^{bk*} + \epsilon_t^{bk}) h_t^{bk} \\
&-\sum_{t\in T}\sum_{b\in B}\sum_{i\in G_b}\phi_{it}^b)\Big\}
\end{aligned} \tag{2–22}$$

$$s.t. \quad \text{Constraints (2–3)} - \text{(2–6)},$$

$$y_{it}^b, u_{it}^b, v_{it}^b \in \{0,1\}, \forall t \in T, \forall b \in B, \forall i \in G_b \tag{2–23}$$

where $\mathcal{X} = \Big\{$

$$\text{Constraints (2–7)} - \text{(2–9)}, \text{(2–18)},$$

$$\sum_{b\in B}\left(\sum_{i\in G_b} x_{it}^b + w_t^b\right) = \sum_{b\in B}\sum_{k\in K} h_t^{bk},\ \forall t \in T \tag{2–24}$$

$$h_t^{bk} \le I_t^{bk},\ \forall t \in T, \forall b \in B, \forall k \in K \tag{2–25}$$

$$-C_{ij} \le \sum_{b\in B} K_{ij}^b \left(\sum_{q\in G_b} x_{qt}^b + w_t^b - \sum_{k\in K} h_t^{bk}\right)$$
$$\le C_{ij}, \forall (i,j) \in \Omega, \forall t \in T \tag{2–26}$$

$$D_{tb}^0 \le \sum_{k\in K} h_t^{bk} \le D_{tb}^M,\ \forall t \in T, \forall b \in B \tag{2–27}$$

$$x_{it}^b \ge 0, h_t^{bk} \ge 0, \tag{2–28}$$
$$\forall t \in T, \forall b \in B, \forall i \in G_b, \forall k \in K\Big\}.$$

Note here that in the above formulation, constraints (2–24) are the reformulations of constraints (2–10), constraints (2–25) are derived from constraints (2–17), constraints (2–26) are derived from constraints (2–11), and constraints (2–27) are the reformulations of constraints (2–12).

2.4 Solution Methodology

For notation brevity, we use matrices and vectors to represent the constraints and variables. The above formulation can be represented as follows:

$$\max_{y,u,v}(\hat{a}^T u + \hat{b}^T v) + \min_{w\in\mathcal{W}} \max_{x,h,\phi\in\chi} (\min_{\epsilon\in\prod} \epsilon^T h + \hat{c}^T h - \hat{e}^T \phi) \tag{2–29}$$

$$s.t.\quad \hat{A}y + \hat{B}u + \hat{C}v \le 0, \tag{2–30}$$

$$y, u, v \in \{0,1\}, \tag{2–31}$$

where

$$\chi = \Big\{\hat{D}x \le \hat{F}y + \hat{g}, \tag{2–32}$$

$$\hat{P}x - \hat{J}\phi \le \hat{K}y, \tag{2–33}$$

$$\hat{R}h \le \hat{q}, \tag{2–34}$$

$$\hat{T}x + \hat{S}h \le \hat{s} + \hat{O}w, \tag{2–35}$$

$$x, h \ge 0\Big\}, \tag{2–36}$$

$$\Pi = \left\{ \hat{U}\epsilon \geq \hat{m} \right\}. \tag{2–37}$$

Constraint (2–30) represents constraints (2–3) - (2–6); constraint (2–32) represents constraints (2–7) - (2–9); constraint (2–33) represents constraints (2–18); constraint (2–34) represents constraints (2–25) and (2–27); constraint (2–35) represents constraints (2–24) and (2–26); constraint (2–36) represents constraints (2–28); constraint (2–37) represents constraints (2–19) and (2–20).

2.4.1 Problem Reformulation

To solve the above formulation, we first dualize the constraints in (2–37) and combine it with the second stage decision variables and constraints. We obtain the following two-stage formulation:

$$\max_{y,u,v}(\hat{a}^T u + \hat{b}^T v) + \min_{w \in \mathcal{W}} \max_{x,h,\phi,\theta \in \bar{\chi}} (\hat{m}^T \theta + \hat{c}^T h - \hat{e}^T \phi) \tag{2–38}$$

$$s.t. \quad \hat{A}y + \hat{B}u + \hat{C}v \leq 0, \tag{2–39}$$

$$y, u, v \in \{0, 1\}, \tag{2–40}$$

where

$$\bar{\chi} = \chi \cap \left\{ \hat{U}^T \theta - h \leq 0, \tag{2–41}$$

$$\theta \geq 0 \right\}. \tag{2–42}$$

Due to the special problem structure (for instance, the third stage uncertainty is only involved in the objective function), taking the dual formulation does not generate nonlinear terms for the resulting two-stage robust optimization problem. Then, we can dualize the remaining constraints in $\bar{\chi}$, and transform the second-stage problem as follows:

$$\begin{aligned} \omega(y) = \quad & \min_{w \in \mathcal{W}, \gamma, \lambda, \tau, \mu, \eta} (\hat{F}y + \hat{g})^T \gamma + (\hat{K}y)^T \lambda + \hat{q}^T \tau \\ & + (\hat{s} + \hat{O}w)^T \mu \end{aligned} \tag{2–43}$$

$$s.t. \quad \hat{D}^T\gamma + \hat{P}^T\lambda + \hat{T}^T\mu \geq 0, \tag{2–44}$$

$$\hat{R}^T\tau + \hat{S}^T\mu - \eta \geq \hat{c}, \tag{2–45}$$

$$\hat{J}^T\lambda = \hat{e}, \tag{2–46}$$

$$\hat{U}\eta \geq \hat{m}, \tag{2–47}$$

$$\gamma, \lambda, \tau, \mu, \eta \geq 0, \tag{2–48}$$

where $\gamma, \lambda, \tau, \mu, \eta$ are dual variables for constraints (2–32), (2–33), (2–34), (2–35), and (2–41) respectively.

In the above formulation, we have the bilinear term: $w^T\hat{O}^T\mu$. Let $\hat{O}^T\mu = \sigma$, and by using the uncertainty set $\mathcal{W}$, we have

$$w^T\sigma = \sum_{t\in T}\sum_{b\in B}\sigma_t^b w_t^b \tag{2–49}$$

$$= \sum_{t\in T}\sum_{b\in B}\sigma_t^b(W_t^{b*} + z_t^{b+}W_t^{b+} - z_t^{b-}W_t^{b-}) \tag{2–50}$$

$$= \sum_{t\in T}\sum_{b\in B}(\sigma_t^b W_t^{b*} + \sigma_t^{b+}W_t^{b+} + \sigma_t^{b-}W_t^{b-}) \tag{2–51}$$

$$s.t. \quad \sigma_t^b = (\hat{O}^T\mu)_t^b, \ \forall t \in T, \forall b \in B \tag{2–52}$$

$$\sigma_t^{b+} \geq -Mz_t^{b+}, \ \forall t \in T, \forall b \in B \tag{2–53}$$

$$\sigma_t^{b+} \geq \sigma_t^b - M(1 - z_t^{b+}), \ \forall t \in T, \forall b \in B \tag{2–54}$$

$$\sigma_t^{b-} \geq -Mz_t^{b-}, \forall t \in T, \ \forall b \in B \tag{2–55}$$

$$\sigma_t^{b-} \geq -\sigma_t^b - M(1 - z_t^{b-}), \ \forall t \in T, \forall b \in B \tag{2–56}$$

$$\sum_{t\in T}(z_t^{b+} + z_t^{b-}) \leq \pi^b, \ \forall b \in B. \tag{2–57}$$

Now we can replace the bilinear term $w^T\hat{O}^T\mu$ by (2–51) and add constraints (2–52) to (2–57) to remove the bilinear term.

2.4.2 Benders' Decomposition

We can use the Benders' decomposition algorithm to solve the above three-stage robust optimization problem. We denote ϑ as the second-stage optimal objective value

and then consider the following master problem. By adding feasibility and optimality cuts, we can solve this problem iteratively:

$$\begin{aligned} \max_{y,u,v} \quad & \hat{a}^T u + \hat{b}^T v + \vartheta \\ s.t. \quad & \text{Constraints (2–39) and (2–40),} \\ & \text{Feasibility cuts,} \\ & \text{Optimality cuts.} \end{aligned}$$

2.4.2.1 Feasibility cuts

We use the L-shaped method to generate feasibility cuts. In this case, when we check the feasibility in $\bar{\chi}$, we do not need to consider constraints (2–33), (2–41) and (2–42) since they will not affect the feasibility. The feasibility check problem is shown as follows:

$$\max_{x,h,\kappa_1,\kappa_2} \quad -\sum_{i=1}^{3} \hat{e}^T \kappa_1^i - \sum_{i=1}^{3} \hat{e}^T \kappa_2^i \tag{2–58}$$

$$s.t. \quad \hat{D}x + \kappa_1^1 - \kappa_2^1 \leq \hat{F}y + \hat{g}, \tag{2–59}$$

$$\hat{R}h + \kappa_1^2 - \kappa_2^2 \leq \hat{q}, \tag{2–60}$$

$$\hat{T}x + \hat{S}h + \kappa_1^3 - \kappa_2^3 \leq \hat{s} + \hat{O}w, \tag{2–61}$$

$$\kappa_1 \geq 0, \kappa_2 \geq 0, \tag{2–62}$$

where $\hat{e}$ represents the vector with all components 1. Now we take the dual of the above formulation and replace the nonlinear term $w^T \hat{O}^T \mu$ by using the same scheme:

$$\begin{aligned} \omega^L(y) = \quad & \min_{w \in \mathcal{W}, \hat{\gamma}, \hat{\tau}, \hat{\mu}, \hat{\eta}} (\hat{F}y + \hat{g})^T \hat{\gamma} + \hat{q}^T \hat{\tau} + \hat{s}^T \hat{\mu} \\ & + (W^*)^T \hat{\sigma} + (W^+)^T \hat{\sigma}^+ + (W^-)^T \hat{\sigma}^- \end{aligned} \tag{2–63}$$

$$s.t. \quad \hat{D}^T \hat{\gamma} + \hat{T}^T \hat{\mu} \geq 0, \tag{2–64}$$

$$\hat{R}^T \hat{\tau} + \hat{S}^T \hat{\mu} \geq 0, \tag{2–65}$$

$$\text{Constraints (2–52) to (2–57),}$$

$$\hat{\gamma}, \hat{\tau}, \hat{\mu}, \hat{\sigma}, \hat{\sigma}^{+}, \hat{\sigma}^{-} \in [0, 1], \tag{2–66}$$

where $\hat{\gamma}, \hat{\tau}$, and $\hat{\mu}$ are dual variables for constraints (2–59), (2–60), and (2–61). Decision variables $\hat{\sigma}$, $\hat{\sigma}^{+}$, $\hat{\sigma}^{-}$ are defined as (2–52) to (2–56). Then we can perform the following steps to check feasibility:

(1) If $\omega^{L}(y) = 0$, y is feasible;

(2) If $\omega^{L}(y) < 0$, generate a corresponding feasibility cut $\omega^{L}(y) \geq 0$.

2.4.2.2 Optimality cuts

Assuming in the ith iteration, we solve the master problem and obtain ϑ^{i} and y^{i}. Since ϑ is the second-stage optimal objective value, if we substitute y^{i} into the subproblem and get $\omega(y^{i})$, we should have $\omega(y^{i}) \geq \vartheta^{i}$. If $\omega(y^{i}) < \vartheta^{i}$, we can claim that y^{i} is not an optimal solution and we can generate a corresponding optimality cut as follows:

$$\omega(y) \geq \vartheta.$$

2.5 Case Study

We evaluate the performance of our proposed approach by testing a revised IEEE 118-bus system available online at *http://motor.ece.iit.edu/data*. The system contains 118 buses, 33 generators, and 186 transmission lines. Apart from the original IEEE 118-bus system, we create the 118SW system by adding a single wind farm at bus 10 and create the 118TW by adding three wind farms at three different buses. The operational time horizon is 24 hours, and each time period is set to be one hour. We use a five-piece piecewise linear function to approximate the generation cost function [31] and use ten-piece piece-wise segments [72] to approximate the price-elastic demand curve. The reference point for the demand response curve is $(80, 25)$ [61]. The pattern of the wind power output including its upper and lower bounds is illustrated in Fig. 2-4, based on the statistics from National Renewable Energy Laboratory (NREL).

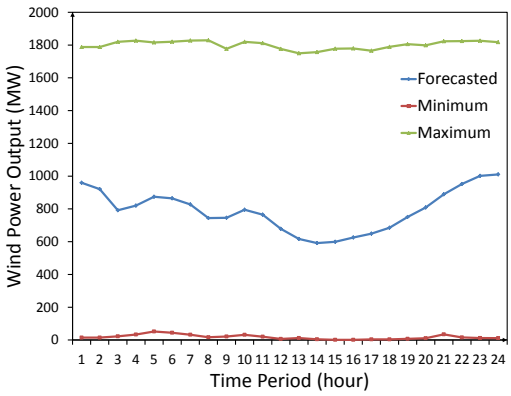

Figure 2-4. Wind power output evolution over time

Finally, all the experiments are implemented using CPLEX 12.1, at Intel Quad Core 2.40GHz with 8GB memory.

2.5.1 Different Demand Response Scenarios

We first consider the effects of different demand response scenarios. To show the effects, we assume the demand response curve is certain and elasticity α_t^b is the same for each b and t (denoted it as α). We set the inelastic demand equal to 80% of the forecasted demand and the demand's upper limit equal to 120% of the forecasted demand. In this case, we test three different elasticity values [72], e.g., $\alpha = -0.5$, $\alpha = -1$, and $\alpha = -2$. We test the 118SW system with four different cardinality budgets π, e.g., $\pi = 2$, $\pi = 4$, $\pi = 6$, and $\pi = 8$. To compare their performance, for each system, we compute the unit load cost (ULC), which is equal to the total cost (i.e., unit commitment cost plus fuel cost) divided by the total demand. We introduce a linear penalty cost function for the unsatisfied demand or transmission capacity/ramp rate limit violations, and the unit penalty cost is set to be 7947/MWh [31]. We report the results of ULC, the social welfare, and the CPU time in Table 2-1.

From Table 2-1 we can observe that ULC has a tendency to decrease as the elasticity value increases, and the social welfare has a tendency to increase as the elasticity value increases. This indicates that, for this case study, introducing demand response can help reduce the unit load cost and increase the social welfare.

Table 2-1. Different price-elastic demand curve scenarios

π	α	ULC	Social Welfare	Time(s)
	-0.5	12.205	1.169e+07	601
2	-1	12.239	7.032e+06	579
	-2	12.270	3.276e+06	598
	-0.5	12.339	1.167e+07	729
4	-1	12.368	7.018e+06	713
	-2	12.409	3.261e+06	706
	-0.5	12.416	1.166e+07	991
6	-1	12.445	7.010e+06	1005
	-2	12.485	3.253e+06	985
	-0.5	12.535	1.165e+07	1373
8	-1	12.553	6.998e+06	1318
	-2	12.603	3.239e+06	1296

2.5.2 Wind Power Output Uncertainty

In this subsection, we test the 118SW system under various wind cardinality budgets (e.g., different π values). We test the system under two different settings. Under the first setting (denoted as CD), we derive robust unit commitment decisions with the consideration of wind power output uncertainty, but without demand response. We set the demand to be the forecasted demand. Under the second setting (denoted as DR), we derive robust unit commitment decisions with the consideration of wind power output uncertainty and deterministic demand response. That is, to show the effectiveness of the demand response, we consider the case in which only wind power output uncertainty is considered and the demand response curve is assumed to be certain. For the demand response curve, we set the elasticity value $\alpha = -1$. We report the results of ULC, the CPU time, and the total number of Benders' cuts (also referred to the number of iterations) in Table 2-2.

From the results reported in Table 2-2 we can observe that, under the worst-case wind power output scenario, the case with demand response has less unit load cost than the case without demand response given the same wind cardinality budget π.

Table 2-2. The comparison of two settings

π	System	ULC	Time(s)	Cuts
2	CD	12.518	320	11
	DR	12.239	579	10
4	CD	12.624	413	15
	DR	12.368	713	11
6	CD	12.722	893	14
	DR	12.455	1005	13
8	CD	12.814	1126	14
	DR	12.553	1318	14

2.5.3 Wind Power Output and Demand Response Uncertainties

In this subsection, we analyze the case, denoted as UDR, considering both wind power output and demand response uncertainties. Besides the uncertainty settings for the wind power output described in Subsection 2.5.2, we test two different deviation values for the price-elastic demand curve, e.g., $\epsilon_t^{bk} = 5$ and $10, \forall t \in T, \forall b \in B, \forall k \in K$. In addition, we set $\varpi_t^b = 0.5 \sum_{k \in K} \epsilon_t^{bk}, \forall t \in T, \forall b \in B$. We also compare these settings with the case in which $\epsilon_t^{bk} = 0, \forall t \in T, \forall b \in B, \forall k \in K$, i.e., the deterministic DR case. The results of ULC, the social welfare (denoted as S.W.), and the consumer surplus and the supplier surplus as defined in Section 2.3 are shown in Table 2-3.

From Table 2-3, we can observe the following:

1. Given a deviation ϵ but different wind cardinality budgets π, under the worst-case scenario when we increase the value of π, ULC increases and the social welfare, both consumer surplus and supplier surplus, decreases. This is due to the need to provide more generation to guarantee that the supply meets the demand.

2. For a fixed wind cardinality budget π but with different deviations ϵ, the social welfare decreases when the system has more demand response uncertainty. This is because we consider the social welfare under the worst-case scenario. When we have more demand response uncertainty, the algorithm gives us a more conservative solution. We also observe that ULC and the supplier surplus decreases when the deviation value increases. But the consumer surplus doesn't change too much as the deviation value increases. The reason for this is that when the demand response curve has more uncertainty, the curve shifts to achieve a

Table 2-3. The uncertain demand response case

π	ϵ	ULC	S.W.(e+6)	Consumer (e+6)	Supplier(e+6)
2	0	12.239	7.03246	4.09943	2.93303
	5	12.205	6.51834	4.10654	2.4118
	10	12.002	6.01348	4.11373	1.89975
4	0	12.368	7.01773	4.08626	2.93147
	5	12.339	6.50367	4.09272	2.41095
	10	12.142	5.99897	4.10004	1.89893
6	0	12.445	7.00967	4.07826	2.93141
	5	12.416	6.49478	4.08475	2.41003
	10	12.221	5.99113	4.09229	1.89884
8	0	12.553	6.99754	4.06665	2.93089
	5	12.535	6.4824	4.0725	2.4099
	10	12.349	5.97854	4.07975	1.89879

smaller demand equilibrium point, which corresponds to smaller total system cost and ULC.

2.5.4 $118\,TW$ System

We report the experiment results for the $118\,TW$ system in this subsection and analyze the impact of the demand response on distributed wind resources. In total, there are three wind farms at three different buses: buses 10, 26, and 32. The output of each wind farm follows the distribution given by Fig. 2-4. We test system CD and system DR as defined in Section IV-B under different wind cardinality budget π, and report ULC, the CPU time, and the total number of Benders' cuts in Table 2-4:

Table 2-4. The comparison of two systems with multiple wind sources

π	System	ULC	Time(s)	Cuts
1	CD	12.9906	993	12
	DR	12.8021	1272	9
2	CD	13.0588	1533	14
	DR	12.8335	2086	9
4	CD	13.1336	2875	13
	DR	12.9045	3408	7

From the results in Table 2-4, we can observe that, under the worst-case wind power output scenario, with multiple wind resources, the case with demand response still has smaller unit load cost than the case without demand response given the same wind cardinality budget π. However, the CPU time increases dramatically as the number of wind farms increases.

2.6 Summary

In this chapter, we developed a robust optimization approach for unit commitment to maximize social welfare under worst-case wind power output and demand response scenarios. An uncertain price-elastic demand curve is used to model consumers' response to price signals. Benders' decomposition is applied to solve the problem. Our final computational results on an IEEE 118-bus system verify that our proposed approach can accommodate both wind power and demand response uncertainties, and demand response can help accommodate wind power output uncertainty by lowering the unit load cost.

Note here that in this chapter, we only consider the wind power output and demand response uncertainties. However, our approach can be applied to other system uncertainties. For instance, in our model setting, we assume the inelastic demand is certain. In practice, the inelastic demand can be uncertain. For this case, we can regard this part of inelastic demand as negative supply and combine it with wind power output uncertainty to construct our uncertainty set. We can also separate it from wind power output uncertainty and build a separate uncertainty set as we did for the wind power output uncertainty case in Section 2.3. The same decomposition approach described in this chapter can be applied to solve the problem. Meanwhile, for the current electricity markets, ancillary service (regulation-up, regulation-down, spin and non-spin) is regarded as a more common way to accommodate wind uncertainty, as compared to demand response.

CHAPTER 3
UNIFIED STOCHASTIC AND ROBUST UNIT COMMITMENT

3.1 Problem Description and Literature Review

As mentioned in Section 1, there are broadly two branches of two-stage optimization approaches to deal with the unit commitment problem under uncertainties, namely two-stage stochastic optimization approach and two-stage robust optimization approach. For the two-stage stochastic optimization approach, the uncertain parameter is generally captured by a series of scenarios, and the objective is to minimize the expected total cost or maximize the expected total social welfare over all the scenarios. Recently, significant work has been done by using the stochastic optimization models to solve the unit commitment problem under uncertainties, from both load side and supply side [70–72]. Although the two-stage stochastic optimization approach is an effective approach to solve the UC problem, it has some practical limitations. First, since the uncertain parameter is often assumed to follow a probability distribution, which is estimated by the historical realizations of the uncertain parameter, the accuracy of probability distribution estimation is critical to determine the optimal UC decisions. However, in practice, it is usually difficult to estimate the distribution of the uncertain parameter precisely. Second, to guarantee a reliability unit commitment run, a large number of scenarios should be generated, which always brings challenges for the stochastic optimization approach to deal with large-scaled instances. However, if the number of selected sceneries is reduced, the feasibility issues may occur. That is, the day-ahead unit commitment decision might not be feasible for some scenarios which are not considered. Therefore, on the other hand, robust optimization approaches have drawn more attention recently for system operators to ensure the system robustness

and make the day-ahead unit commitment feasible for most outcomes of the real time uncertain parameter.

For the robust optimization approach, the uncertain parameter is described within a given deterministic uncertainty set and the objective is to minimize the total cost under the worst-case scenario. The first-stage decision is the unit commitment decision that is made before the realization of the worst-case scenario, and the second-stage decision is the economic dispatch decision which is made after the worst-case scenario is realized. The advantage of the robust optimization approach is that it does not require the distribution information of the input uncertain parameter, as the stochastic optimization approaches do. Instead, as long as the information is sufficient to generate the deterministic uncertainty set, the day-ahead unit commitment decision can be obtained to be feasible for most outcomes of the real time uncertain problem parameter. Recently, two-stage robust unit commitment models are proposed and Benders' decomposition algorithm is utilized to ensure system robustness under uncertainties [7, 32, 78]. Since robust optimization achieves the optimal solution with the consideration of worst-case scenario, this approach always faces the challenges on its over conservatism, due to its objective function of minimizing the worst-case cost, because the worst case happens rarely.

To address the shortages of stochastic optimization approach and robust optimization approach, in this chapter, we propose an innovative unified stochastic and robust unit commitment model to take advantage of both stochastic and robust optimization approaches. In our model, we put the weights for the expected total cost and the worst-case cost respectively in the objective function. This approach allows the system operators to decide the weight for each objective based on their preferences. In addition, if the weight in the objective function for the robust optimization part is zero, then our model turns out to be a two-stage stochastic optimization problem with additional

constraints generated based on the robust optimization approach. The main contribution of our proposed innovative approach can be summarized as follows:

(1) Our proposed approach takes the advantages of both the stochastic and robust optimization approaches. Our approach can provide a day-ahead unit commitment decision that can lead to a minimum expected total cost while ensuring the system robustness.

(2) Our proposed approach can generate a less conservative solution as compared to the two-stage robust optimization approach, and a more robust solution as compared to the two-stage stochastic optimization approach.

(3) Our proposed approach can be implemented in a single Benders' decomposition framework. The computational time can be controlled by the system operators. Meanwhile, the system operators can also adjust the weights in the objective function, based on their preferences on stochastic and/or robust optimization approaches.

The remainder of this chapter is organized as follows. Section 3.3 describes our unified stochastic and robust unit commitment model. In Section 3.4, we develop a Benders' decomposition algorithm to solve the problem. In the proposed algorithm, we apply both feasibility and optimality cuts for the stochastic and robust optimization parts respectively. All these cuts are added into the master problem, which provides the day-ahead unit commitment decision. Section 3.5 provides and analyzes the computational experiments through several case studies. Finally, Section 3.6 summarizes our research.

3.2 Nomenclature

A. Sets and Parameters

B Index set of all buses.

Ω Index set of transmission lines linking two buses.

G_b Set of thermal generators at bus b.

T Time horizon (e.g., 24 hours).

SU_i^b Start-up cost of thermal generator i at bus b.

SD_i^b Shut-down cost of thermal generator i at bus b.

$F_i(.)$ Fuel cost of thermal generator i.

MU_i^b Minimum up-time for thermal generator i at bus b.

MD_i^b Minimum down-time for thermal generator i at bus b.

RU_i^b Ramp-up rate limit for thermal generator i at bus b.

RD_i^b Ramp-down rate limit for thermal generator i at bus b.

L_i^b Lower bound of electricity generated by thermal generator i at bus b.

U_i^b Upper bound of electricity generated by thermal generator i at bus b.

C_{ij} Capacity for the transmission line linking bus i and bus j.

K_{ij}^b Line flow distribution factor for the transmission line linking bus i and bus j, due to the net injection at bus b.

π_t^b The weight of the load at bus b in time t.

$\bar{\pi}^t$ The budget parameter to describe the uncertainty set for the total load in time t.

$\bar{\pi}$ The budget parameter to describe the uncertainty set for the total load for the whole operational horizon.

α The weight for the expected total generation cost in the objective function.

D_t^{b+} The upper bound of the load at bus b in time t.

D_t^{b-} The lower bound of the load at bus b in time t.

D_t^{b*} The forecasted load at bus b in time t.

d_t^b A random parameter representing the load at bus b in time t for the robust optimization part.

$d_t^b(\xi)$ The load at bus b in time t corresponding to scenario ξ for the stochastic optimization part.

γ_{it}^{jb} The intercept of the jth segment line for the generation cost for generator i at bus b in time t.

β_{it}^{jb} The slope of the jth segment line for the generation cost for generator i at bus b in time t.

B. First-stage Variables

y_{it}^{b} Binary decision variable: "1" if thermal generator i at bus b is on in time t; "0" otherwise.

u_{it}^{b} Binary decision variable: "1" if thermal generator i at bus b is started up in time t; "0" otherwise.

v_{it}^{b} Binary decision variable: "1" if thermal generator i at bus b is shut down in time t; "0" otherwise.

C. Second-stage Variables

$q_{it}^{b}(\xi)$ Electricity generation amount by thermal generator i at bus b in time t corresponding to scenario ξ for the stochastic optimization part.

x_{it}^{b} Electricity generation amount by thermal generator i at bus b in time t for the robust optimization part.

$\phi_{it}^{b}(\xi)$ Auxiliary variable for the stochastic optimization part representing the fuel cost of thermal generator i at bus b in time t corresponding to scenario ξ.

ϕ_{it}^{b} Auxiliary variable for the robust optimization part representing the fuel cost of thermal generator i at bus b in time t.

3.3 Mathematical Formulation

In this section, we develop a two-stage unit commitment formulation considering both the expected total generation cost and the worst-case scenario generation cost. The first stage is to determine the day-ahead unit commitment decision that includes turn-on/turn-off decisions of thermal generators by satisfying unit commitment physical constraints. The second stage contains the decisions on the economic dispatch for the thermal generators under each scenario for the stochastic optimization part and the worst-case scenario for the robust optimization part. In our model, a parameter $\alpha \in [0, 1]$ is introduced to represent the weight of the expected total generation cost and accordingly $1 - \alpha$ represents the weight of the worst-case generation cost. Especially, for the case $\alpha = 1$, only the expected total generation cost is considered in the objective function and for the case $\alpha = 0$, only the worst-case generation cost is considered. The

detailed formulation is described as follows:

$$\min \sum_{t=1}^{T} \sum_{b \in B} \sum_{i \in G_b} (SU_i^b u_{it}^b + SD_i^b v_{it}^b) + \alpha E[\mathcal{Q}(y, u, v, \xi)] +$$

$$(1-\alpha) \max_{d \in \mathcal{D}} \min_{x \in \chi(y,u,v,d)} \sum_{t=1}^{T} \sum_{b \in B} \sum_{i \in G_b} F_i(x_{it}^b) \tag{3–1}$$

$s.t.$

$$-y_{i(t-1)}^b + y_{it}^b - y_{ik}^b \leq 0,$$

$$\forall k : 1 \leq k - (t-1) \leq MU_i^b, \forall i \in G_b, \forall b \in B, \forall t \tag{3–2}$$

$$y_{i(t-1)}^b - y_{it}^b + y_{ik}^b \leq 1,$$

$$\forall k : 1 \leq k - (t-1) \leq MD_i^b, \forall i \in G_b, \forall b \in B, \forall t \tag{3–3}$$

$$-y_{i(t-1)}^b + y_{it}^b - u_{it}^b \leq 0, \quad \forall i \in G_b, \forall b \in B, \forall t \tag{3–4}$$

$$y_{i(t-1)}^b - y_{it}^b - v_{it}^b \leq 0, \quad \forall i \in G_b, \forall b \in B, \forall t \tag{3–5}$$

$$y_{it}^b, u_{it}^b, v_{it}^b \in \{0, 1\}, \quad \forall i \in G_b, \forall b \in B, \forall t, \tag{3–6}$$

where $\mathcal{Q}(y, u, v, \xi)$ is equal to

$$\min \sum_{t=1}^{T} \sum_{b \in B} \sum_{i \in G_b} F_i(q_{it}^b(\xi)) \tag{3–7}$$

$s.t.$

$$L_i^b y_{it}^b \leq q_{it}^b(\xi) \leq U_i^b y_{it}^b, \ \forall i \in G_b, \forall b \in B, \forall t \tag{3–8}$$

$$q_{it}^b(\xi) - q_{i(t-1)}^b(\xi) \leq (2 - y_{i(t-1)}^b - y_{it}^b) L_i^b +$$

$$(1 + y_{i(t-1)}^b - y_{it}^b) UR_i^b, \ \forall i \in G_b, \forall b \in B, \forall t \tag{3–9}$$

$$q_{i(t-1)}^b(\xi) - q_{it}^b(\xi) \leq (2 - y_{i(t-1)}^b - y_{it}^b) L_i^b +$$

$$(1 - y_{i(t-1)}^b + y_{it}^b) DR_i^b, \ \forall i \in G_b, \forall b \in B, \forall t \tag{3–10}$$

$$\sum_{b \in B} \sum_{i \in G_b} q_{it}^b(\xi) = \sum_{b \in B} d_{bt}(\xi), \forall t \tag{3–11}$$

$$-C_{ij} \leq \sum_{b \in B} K_{ij}^b (\sum_{r \in G_b} q_{rt}^b(\xi) - d_{bt}(\xi)) \leq C_{ij},$$

$$\forall (i, j) \in \Omega, \forall t \tag{3–12}$$

and

$$\chi(y, u, v, d) = \Big\{ x :$$

$$L_i^b y_{it}^b \le x_{it}^b \le U_i^b y_{it}^b, \ \forall i \in G_b, \forall b \in B, \forall t \tag{3–13}$$

$$x_{it}^b - x_{i(t-1)}^b \le (2 - y_{i(t-1)}^b - y_{it}^b) L_i^b +$$
$$(1 + y_{i(t-1)}^b - y_{it}^b) UR_i^b, \ \forall i \in G_b, \forall b \in B, \forall t \tag{3–14}$$

$$x_{i(t-1)}^b - x_{it}^b \le (2 - y_{i(t-1)}^b - y_{it}^b) L_i^b +$$
$$(1 - y_{i(t-1)}^b + y_{it}^b) DR_i^b, \ \forall i \in G_b, \forall b \in B, \forall t \tag{3–15}$$

$$\sum_{b \in B} \sum_{i \in G_b} x_{it}^b = \sum_{b \in B} d_{bt}, \forall t \tag{3–16}$$

$$-C_{ij} \le \sum_{b \in B} K_{ij}^b \Big(\sum_{r \in G_b} x_{rt}^b - d_{bt} \Big) \le C_{ij},$$
$$\forall (i, j) \in \Omega, \forall t \Big\}. \tag{3–17}$$

In the above formulation, we denote $F_i(.)$ as the generation cost function of generator i. The objective function (3–1) is composed of the unit commitment cost in the first stage, and both the expected economic dispatch cost and the worst-case economic dispatch cost in the second stage. Constraints (3–2) and (3–3) represent each unit's minimum up-time and minimum down-time restrictions respectively. Constraints (3–4) and (3–5) indicate the start-up and shut-down operations for each unit. Constraints (3–8) and (3–13) enforce the upper and lower limits of the power generation amount of each unit. The ramping up constraints (3–9) and (3–14) require the first-hour minimum generation restriction (e.g., L_i^b) as described in [27] and limit the maximum increment of the power generation amount of each unit between two adjacent periods when the generator is on. Similarly, ramping down constraints (3–10) and (3–15) require the last-hour minimum generation restriction (e.g., L_i^b) and enforce the maximum decrement of the power generation amount of each unit between two adjacent periods when the generator is on. Constraints (3–11) and (3–16) ensure load balance and constraints (3–12) and (3–17) represent the transmission capacity constraints.

3.4 Decomposition Algorithms and Solution Framework

3.4.1 Scenario Generation

We use Monte Carlo simulation to generate scenarios for the uncertain load. We assume that the load follows a multivariate normal distribution $N(D, \Sigma)$ with its predicted value D and volatility matrix Σ. We can run Monte Carlo simulation to generate N scenarios each with the same probability $1/N$. After generating scenarios, we can replace the second-stage expected total cost objective term by:

$$\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{b\in B}\sum_{i\in G_b} F_i(q_{it}^b(\xi^n)). \tag{3–18}$$

3.4.2 Linearizing $F_i(.)$

The generation cost $F_i(.)$ is usually expressed as a quadratic function, for which we use a J-piece piecewise linear function to approximate. For instance, we have

$$\begin{aligned} &\phi_{it}^b(\xi^n) \geq \gamma_{it}^{jb} y_{it}^b + \beta_{it}^{jb} q_{it}^b(\xi^n), \\ &\forall t \in T, \forall b \in B, \forall i \in G_b, \forall j = 1, \cdots, J, \forall n = 1, \cdots, N \end{aligned} \tag{3–19}$$

for the stochastic optimization part, and similarly, we have

$$\begin{aligned} &\phi_{it}^b \geq \gamma_{it}^{jb} y_{it}^b + \beta_{it}^{jb} x_{it}^b, \\ &\forall t \in T, \forall b \in B, \forall i \in G_b, \forall j = 1, \cdots, J \end{aligned} \tag{3–20}$$

for the robust optimization part.

3.4.3 The Uncertainty Set of the Load

To generate the uncertainty set for the robust optimization part, we assume the load for each time period t at each bus b is between a lower bound D_t^{b-} and an upper bound D_t^{b+}, which can be decided by the 5th and 95th percentiles of the random load output. In addition, we assume for each given time period t, the summation of the weighted loads at all buses is bounded above by $\bar{\pi}^t$, and the summation of the weighted loads within the

whole operational horizon is bounded above by $\bar{\pi}$. Accordingly, the uncertainty set can be described as follows:

$$\mathcal{D} := \Bigg\{ d \in \mathcal{R}^{|B|\times|T|} : D_t^{b-} \leq d_t^b \leq D_t^{b+}, \forall t, \forall b \tag{3–21}$$

$$\sum_{b\in B} \pi_b^t d_b^t \leq \bar{\pi}^t, \forall t \tag{3–22}$$

$$\sum_{t=1}^{T} \sum_{b\in B} \pi_b^t d_b^t \leq \bar{\pi} \Bigg\}. \tag{3–23}$$

3.4.4 Abstract Formulation

For notation brevity, we use matrices and vectors to represent the constraints and variables. For example, we use **e** to represent the vector with all components equal to 1. The mathematical model can be abstracted as follows:

$$\min_{y,u,v} (\mathbf{a}^T u + \mathbf{b}^T v) + \alpha \frac{1}{N} \sum_{n=1}^{N} \mathbf{e}^T \phi(\xi^n) + (1-\alpha) \max_{d\in\mathcal{D}} \min_{x,\phi} \mathbf{e}^T \phi$$

$$s.t.\ \ \mathbf{A}y + \mathbf{B}u + \mathbf{C}v \geq \mathbf{r}, \tag{3–24}$$

$$\mathbf{F}y - \mathbf{D}q(\xi^n) \leq \mathbf{g},\ \ n = 1, \cdots, N \tag{3–25}$$

$$\mathbf{K}y - \mathbf{P}q(\xi^n) - \mathbf{J}\phi(\xi^n) \leq 0,\ \ n = 1, \cdots, N \tag{3–26}$$

$$\mathbf{T}q(\xi^n) \geq \mathbf{S}d(\xi^n) + \mathbf{s},\ \ n = 1, \cdots, N \tag{3–27}$$

$$\mathbf{F}y - \mathbf{D}x(d) \leq \mathbf{g}, \tag{3–28}$$

$$\mathbf{K}y - \mathbf{P}x(d) - \mathbf{J}\phi(d) \leq 0, \tag{3–29}$$

$$\mathbf{T}x(d) \geq \mathbf{S}d + \mathbf{s}, \tag{3–30}$$

$$y, u, v \in \{0,1\}, x(d), q(\xi^n) \geq 0, \phi(d), \phi(\xi^n) \text{ free}, \forall n \tag{3–31}$$

where

$$\mathcal{D} = \Big\{ d \in \mathcal{R}^{|B|\times|T|} : \mathbf{d}^- \leq d \leq \mathbf{d}^+, \mathbf{U}^T d \leq \mathbf{z} \Big\}. \tag{3–32}$$

Constraint (3–24) represents constraints (3–2) - (3–5); constraint (3–25) represents constraints (3–8) - (3–10); constraint (3–26) represents constraint (3–19); constraint

(3–27) represents constraints (3–11) and (3–12). Constraint (3–28) represents constraints (3–13) - (3–15); constraint (3–29) represents constraint (3–20); constraint (3–30) represents constraints (3–16) and (3–17).

3.4.5 Benders' Decomposition Algorithm

We can use the Benders' decomposition algorithm to solve the above problem. First, for each scenario ξ^n, $n = 1, \cdots, N$, we dualize the constraints (3–25) to (3–27) and obtain the following dual formulation for the second-stage economic dispatch for the stochastic part:

$$\begin{aligned}
\psi^{S_n}(y) = \max_{\gamma^n, \lambda^n, \mu^n} & (\mathbf{F}y - \mathbf{g})^T \gamma^n + (\mathbf{K}y)^T \lambda^n + (\mathbf{s} + \mathbf{S}d(\xi^n))^T \mu^n \\
s.t. \quad & \mathbf{D}^T \gamma^n + \mathbf{P}^T \lambda^n + \mathbf{T}^T \mu^n \leq 0, && \text{(3–33)} \\
& \mathbf{J}^T \lambda^n = \mathbf{e}, && \text{(3–34)} \\
& \gamma^n, \lambda^n, \mu^n \geq 0, && \text{(3–35)}
\end{aligned}$$

where $\gamma^n, \lambda^n, \mu^n$ are dual variables corresponding to the scenario n for constraints (3–25), (3–26) and (3–27) respectively.

Similarly, we dualize the constraints (3–28) to (3–30) and obtain the following dual formulation for the second-stage economic dispatch for the robust part:

$$\begin{aligned}
\psi^R(y) = \max_{d \in \mathcal{D}, \gamma, \lambda, \mu} & (\mathbf{F}y - \mathbf{g})^T \gamma + (\mathbf{K}y)^T \lambda + (\mathbf{s} + \mathbf{S}d)^T \mu \\
s.t. \quad & \mathbf{D}^T \gamma + \mathbf{P}^T \lambda + \mathbf{T}^T \mu \leq 0, && \text{(3–36)} \\
& \mathbf{J}^T \lambda = \mathbf{e}, && \text{(3–37)} \\
& \gamma, \lambda, \mu \geq 0, && \text{(3–38)}
\end{aligned}$$

where γ, λ, μ are dual variables for constraints (3–28), (3–29) and (3–30) respectively.

We denote θ^n as the second-stage optimal economic dispatch cost corresponding to scenario n, $n = 1, \cdots, N$ and $\bar{\theta}$ as the second-stage optimal economic dispatch cost under the worst-case scenario. Then the master problem can be described as follows,

and the problem can be solved by adding feasibility and optimality cuts iteratively:

$$\begin{aligned} \min_{y,u,v\in\{0,1\}} & (\mathbf{a}^T u + \mathbf{b}^T v) + \alpha \frac{1}{N} \sum_{n=1}^{N} \theta^n + (1-\alpha)\bar{\theta} \\ s.t. \quad & \mathbf{A}y + \mathbf{B}u + \mathbf{C}v \geq \mathbf{r}, \\ & \text{Feasibility cuts}, \\ & \text{Optimality cuts}. \end{aligned}$$

3.4.6 Benders' Cuts for the Stochastic Optimization Part

3.4.6.1 Feasibility cuts

We use the L-shaped method to generate feasibility cuts. In this case, we don't need to consider constraints (3–26) since we can always choose $\phi(\xi^n)$ to make these constraints satisfied. Thus, they will not affect the feasibility. For constraints (3–25) and (3–27) in scenario ξ^n, $n = 1, \cdots, N$, the feasibility check problem is shown as follows:

$$\min_{q(\xi^n),\kappa} \quad \sum_{j=1}^{4} \mathbf{e}^T \kappa^j \tag{3–39}$$

$$s.t. \quad \mathbf{D}q(\xi^n) + \kappa^1 - \kappa^2 \geq \mathbf{F}y - \mathbf{g}, \tag{3–40}$$

$$\mathbf{T}q(\xi^n) + \kappa^3 - \kappa^4 \geq \mathbf{S}d(\xi^n) + \mathbf{s}, \tag{3–41}$$

$$q(\xi^n) \geq 0, \kappa^j \geq 0, j = 1, \cdots, 4. \tag{3–42}$$

The dual of the above formulation can be described as follows:

$$\omega^{S_n}(y) = \max_{\hat{\gamma}^n,\hat{\mu}^n} (\mathbf{F}y - \mathbf{g})^T \hat{\gamma}^n + (\mathbf{S}d(\xi^n) + \mathbf{s})^T \hat{\mu}^n \tag{3–43}$$

$$s.t. \quad \mathbf{D}^T \hat{\gamma}^n + \mathbf{T}^T \hat{\mu}^n \leq 0, \tag{3–44}$$

$$\hat{\gamma}^n, \hat{\mu}^n \in [0, 1], \tag{3–45}$$

where $\hat{\gamma}^n$ and $\hat{\mu}^n$ are dual variables corresponding to the nth scenario for constraints (3–40) and (3–41) respectively. Then we can perform the following steps to check feasibility:

(1) If $\omega^{S_n}(y) = 0$, y is feasible for the nth scenario;

(2) If $\omega^{S_n}(y) > 0$, generate a corresponding feasibility cut $\omega^{S_n}(y) \leq 0$.

3.4.6.2 Optimality cuts

At each iteration, after solving the master problem, we obtain θ^n,$n = 1, \cdots, N$, and y. If we substitute y into the subproblem and get $\psi^{S_n}(y)$, we should have $\psi^{S_n}(y) \leq \theta^n$. If $\psi^{S_n}(y) > \theta^n$, we can claim that y is not an optimal solution and we can generate a corresponding optimality cut as follows:

$$\psi^{S_n}(y) \leq \theta^n.$$

3.4.7 Benders' Cuts for the Robust Optimization Part

After solving the master problem, we apply the bilinear approach to get $\psi^R(y)$. The bilinear approach is proved to converge to optimality with a small gap (less than 0.05%) in a reasonable and much shorter time than the exact separation algorithm does [32]. We solve the following two linear programs iteratively:

$$\psi^{R1}(y, d) = \max_{\gamma,\lambda,\mu} (\mathbf{F}y - \mathbf{g})^T\gamma + (\mathbf{K}y)^T\lambda + (\mathbf{s} + \mathbf{S}d)^T\mu$$

$$\text{(SUB1)} \quad s.t. \ \mathbf{D}^T\gamma + \mathbf{P}^T\lambda + \mathbf{T}^T\mu \leq 0, \tag{3–46}$$

$$\mathbf{J}^T\lambda = \mathbf{e}, \tag{3–47}$$

$$\gamma, \lambda, \mu \geq 0; \tag{3–48}$$

$$\psi^{R2}(y, \gamma, \lambda, \mu) = \max_{d} \mu^T\mathbf{S}d + (\mathbf{F}y - \mathbf{g})^T\gamma + (\mathbf{K}y)^T\lambda + \mathbf{s}^T\mu$$

$$\text{(SUB2)} \quad s.t. \ \mathbf{d}^- \leq d \leq \mathbf{d}^+, \tag{3–49}$$

$$\mathbf{U}^T d \leq \mathbf{z}. \tag{3–50}$$

Next, we discuss the feasibility and optimality cuts based on the proposed approach.

3.4.7.1 Feasibility cuts

If the first-stage solution y is infeasible then we should have $\psi^R(y)$ infeasible or unbounded. Note that $\psi^{R2}(y, \gamma, \lambda, \mu)$ is feasible and bounded from above because the feasible region of d is a polyhedra (e.g., bounded), and the objective function is continuous. Therefore, we only need to check the feasibility of (SUB1). The feasibility check for a given d is the same as that for the stochastic case:

$$\omega^{R1}(y, d) = \max_{\hat{\gamma}, \hat{\mu}} (\mathbf{F}y - \mathbf{g})^T \hat{\gamma} + (\mathbf{s} + \mathbf{S}d)^T \hat{\mu} \tag{3–51}$$

$$(\text{FEA1}) \quad s.t. \ \mathbf{D}^T \hat{\gamma} + \mathbf{T}^T \hat{\mu} \leq 0, \tag{3–52}$$

$$\hat{\gamma}, \hat{\mu} \in [0, 1]. \tag{3–53}$$

Note here that in order to check the feasibility under the worst-case scenario, we need to solve (FEA1) and the following (FEA2) iteratively to find the worst-case load d.

$$\omega^{R2}(y, \hat{\gamma}, \hat{\mu}) = \max_{d} \hat{\mu}^T \mathbf{S}d + (\mathbf{F}y - \mathbf{g})^T \hat{\gamma} + \mathbf{s}^T \hat{\mu} \tag{3–54}$$

$$(\text{FEA2}) \quad s.t. \ \text{Constraints (3–49) and (3–50).}$$

This procedure stops when we have $\omega^{R2}(y, \hat{\gamma}, \hat{\mu}) \leq \omega^{R1}(y, d)$. The detailed algorithm is shown as follows:

(1) Pick an extreme point d in $\mathcal{D}$;

(2) Solve (FEA1), and store the optimal objective value $\omega^{R1}(y, d)$ and the optimal solution $\hat{\gamma}$ and $\hat{\mu}$;

(3) Solve (FEA2), and store the optimal objective value $\omega^{R2}(y, \hat{\gamma}, \hat{\mu})$ and the optimal solution d^*;

(4) If $\omega^{R2}(y, \hat{\gamma}, \hat{\mu}) > \omega^{R1}(y, d)$, let $d = d^*$ and go to Step (2). Otherwise go to Step (5);

(5) If $\omega^{R1}(y, d) = 0$, terminate the feasibility check for the robust part, and go to the feasibility check for the stochastic part. Otherwise, add the feasibility cut $\omega^{R1}(y, d) \leq 0$ to the master problem.

3.4.7.2 Optimality cuts

At each iteration, we solve the master problem and obtain $\bar{\theta}$ and y. Similarly, if $\psi^R(y) > \bar{\theta}$, we can generate a corresponding optimality cut as follows:

$$\psi^R(y) \leq \bar{\theta}.$$

The detailed algorithm to obtain optimality cuts is as follows:

(1) Pick an extreme point d in $\mathcal{D}$;

(2) Solve (SUB1), and store the optimal objective value $\psi^{R1}(y, d)$ and the optimal solution γ and μ;

(3) Solve (SUB2), and store the optimal objective value $\psi^{R2}(y, \gamma, \lambda, \mu)$ and the optimal solution d^*;

(4) If $\psi^{R2}(y, \gamma, \lambda, \mu) > \psi^{R1}(y, d)$, let $d = d^*$ and go to Step (2). Otherwise go to Step (5);

(5) If $\psi^{R2}(y, \gamma, \lambda, \mu) > \bar{\theta}$, generate the corresponding optimality cut $\psi^{R2}(y, \gamma, \lambda, \mu) \leq \bar{\theta}$ to the master problem. Otherwise, terminate the optimality check for the robust part and go to the optimality check for the stochastic part.

Finally, the flow chart of our proposed algorithm is shown in Fig. 3-1.

3.4.8 Special Cases and Discussions

The robust optimization approach helps ensure the robustness of the first-stage unit commitment decision. We can consider a special case in which we only use the constraints provided by the robust optimization approach to guarantee the solution robustness without considering the worst-case economic dispatch cost in the objective function (i.e., $\alpha = 1$). In addition, if we only take a small number of scenarios, then we don't need to use the Benders' decomposition approach to generate cuts for the stochastic part. Instead, we can put the stochastic part in the master problem, and add feasibility cuts by considering the constraints in the robust part. Therefore, the special case can be reformulated as follows:

$$\min_{y,u,v} (\mathbf{a}^T u + \mathbf{b}^T v) + \frac{1}{N} \sum_{n=1}^{N} \mathbf{e}^T \phi(\xi^n)$$

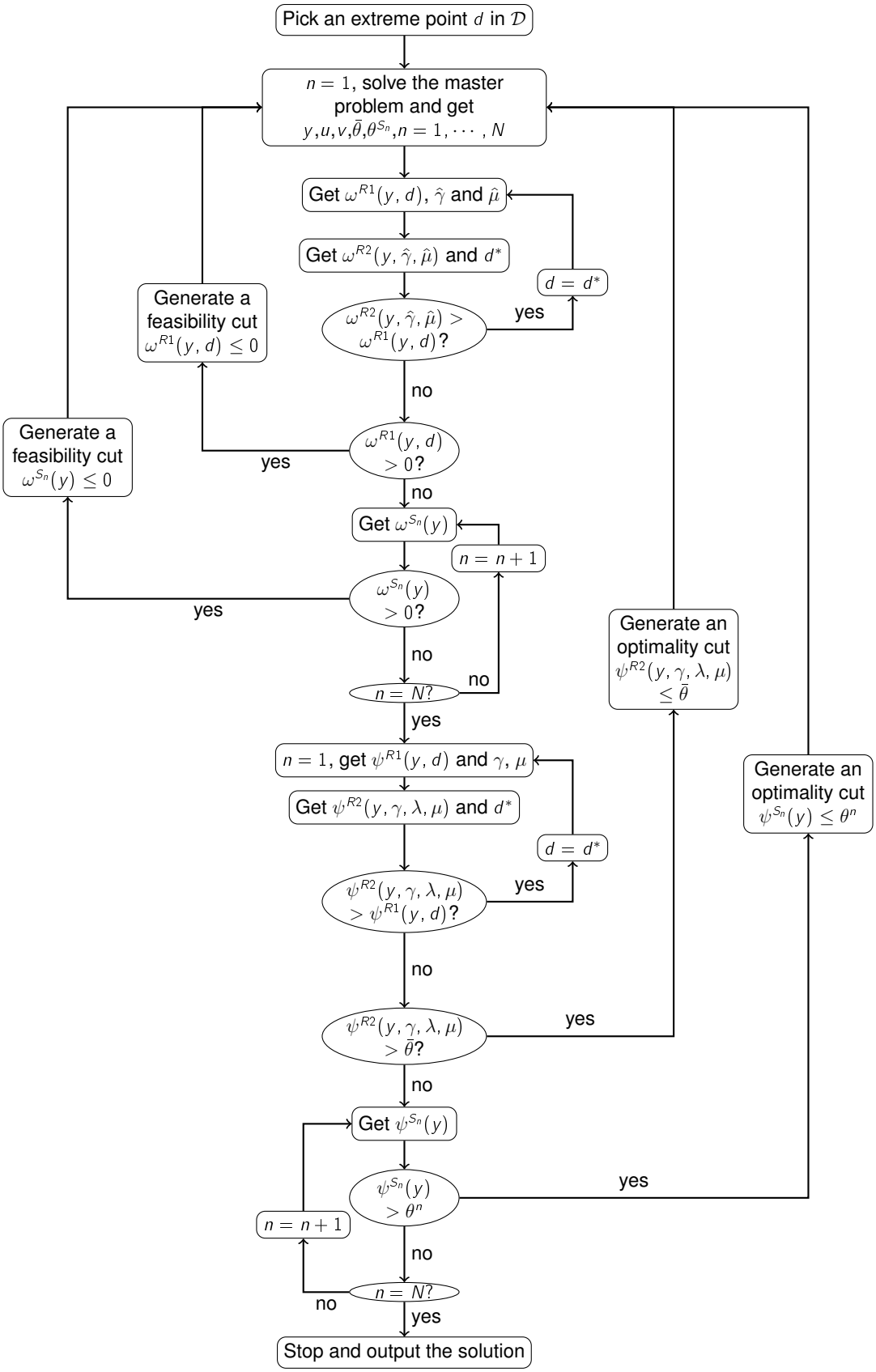

Figure 3-1. Flow chart of the proposed algorithm

$$
\begin{aligned}
s.t.\ & \text{Constraints (3–24) - (3–27)}, \\
& \mathbf{F}y - \mathbf{D}x(d) \leq \mathbf{g}, \mathbf{T}x(d) \geq \mathbf{S}d + \mathbf{s}, \forall d \in \mathcal{D} \\
& y, u, v \in \{0,1\}, x(d) \geq 0, \phi(\xi^n) \text{ free}.
\end{aligned}
$$

For the above formulation, the decision variable $x(d)$ is the auxiliary variable representing the generation amount corresponding to the case in which the load is d. Constraints (3–24) - (3–27) are put in the master problem. We apply the robust feasibility check described in Section 4–2 to find the worst-case load d^*. Then we can add the Benders' feasibility cuts in the form $\omega^{R_1}(y, d^*) \leq 0$ to the master problem.

Note here that in this section, we provide a general framework to solve the unified stochastic and robust unit commitment problem. There are several deviations we can explore for each part in the algorithm. For instance, for the stochastic part, we can choose the Benders' decomposition approach or solve the deterministic equivalent formulation (as above for the special case); for the robust optimality part, we can consider including or not including the worst-case cost in the objective function; for the robust feasibility part, once the worst-case load d^* is detected, we can add the dual cut in the form $\omega^{R_1}(y, d^*) \leq 0$, or a group of primal cuts in the form $\mathbf{F}y - \mathbf{D}x(d^*) \leq \mathbf{g}, \mathbf{T}x(d^*) \geq \mathbf{S}d^* + \mathbf{s}$, similar to the one described in [78].

3.5 Computational Results

In this section, we report experimental results for a modified IEEE 118-bus system, based on the one given online at *http://motor.ece.iit.edu/data*, to show the effectiveness of the proposed approach. The system contains 118 buses, 33 generators, and 186 transmission lines. The operational time interval is 24 hours. In our experiments, we set the feasibility tolerance gap to be 10^{-6} and the optimality tolerance gap $(\psi^{R2} - \bar{\theta})/\bar{\theta}$ to be 10^{-4}. The MIP gap tolerance for the master problem is the CPLEX default gap. We use C++ with CPLEX 12.1 to implement the proposed formulations and algorithms. All experiments are executed on a computer workstation with 4 Intel Cores and 8GB RAM.

In our experiment, we first perform sensitivity analysis of our proposed approach in terms of the effects of the uncertainty set and the objective weight α. Then, we compare the performances of our proposed approach with the stochastic and robust optimization approaches.

3.5.1 Sensitivity Analysis

3.5.1.1 Effect of uncertainty set

For convenience, we normalize the weight parameter $\pi_t^b = 1$. We first let $D_t^{b+} = (1 + \text{Ratio}\%)D_t^{b*}$ and $D_t^{b-} = (1 - \text{Ratio}\%)D_t^{b*}$, $\forall t, \forall b$. Then, for each time t, we let the budget $\bar{\pi}^t = (1 + \text{Budget}\%)\sum_{b\in B} D_t^{b*}$. Finally, we let the overall budget $\bar{\pi} = 0.9\sum_{t=1}^{T}\sum_{b\in B} D_t^{b+}$. In our experiment, we allow the sensitivity analysis parameters Ratio% and Budget% to vary from 0 to 25%. In addition, we use a five-piece piecewise linear function to approximate the generation cost function.

We test the performance of our proposed approach under various Ratio% and Budget% settings. Note here that when Budget% $>$ Ratio%, constraint (3–22) is redundant and accordingly the computational results for these cases are the same as the one in which Budget% $=$Ratio%. In this experiment, we set the sample size N to be 5, and report the optimal objective value, the number of start-ups, and the computational time for each setting in Table 3-1.

Table 3-1. Results under different Ratio% and Budget % settings

Ratio%		Budget%		
		5%	15%	25%
5%	Obj. Val.($)	737,134	737,134	737,134
	# of Start-ups	10	10	10
	Time(sec)	48	48	48
15%	Obj. Val.($)	737,439	738,372	738,272
	# of Start-ups	10	11	11
	Time(sec)	61	163	163
25%	Obj. Val.($)	738,561	739,374	740,175
	# of Start-ups	12	12	13
	Time(sec)	106	392	1193

From results shown in Table 3-1, we can observe that

(1) given an identical Budget%, the objective value and the number of start-ups increase as Ratio% increases, because the problem becomes more conservative, and similarly,

(2) given an identical Ratio%, the objective value and the number of start-ups increase as Budget% increases, because the system allows more load fluctuations.

3.5.1.2 Sensitivity analysis of objective weight α

In this subsection, we test the performance of our proposed approach under various objective weight α settings. Note here that if $\alpha = 0$, only the worst-case total cost is considered and if $\alpha = 1$, only the expected total cost is included. For this experiment, we set Ratio% $= 10\%$ and Budget% $= 10\%$. The computational results are shown in Fig. 4-2.

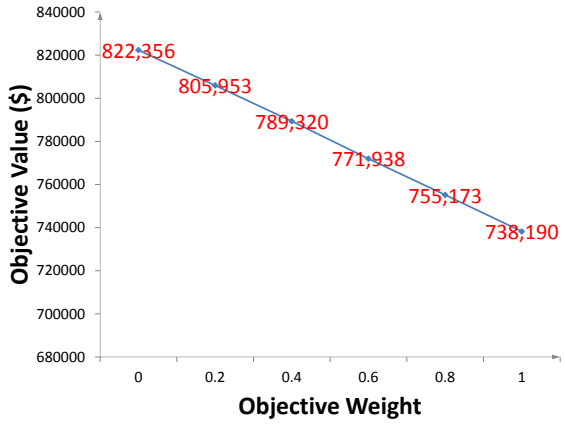

Figure 3-2. The relationship between the objective value and the objective weight

From Fig. 4-2 we can observe that, as α increases, the optimal objective value decreases, because the problem becomes less conservative as the worst-case cost component has a smaller weight.

3.5.2 Proposed Approach vs Stochastic Optimization Approach

In this subsection, we set Ratio% $=$ Budget% and compare the performances of our proposed approach with the traditional two-stage stochastic optimization approach under different Budget% and scenario-size settings. In this experiment, we do not

include the worst-case generation cost in the objective function (i.e., $\alpha = 1$). The total costs corresponding to the traditional two-stage stochastic optimization approach (SO) and our proposed unified stochastic and robust optimization approach (SR) are obtained by the following two steps:

(1) Solve the problem by using the SO and SR approaches respectively, and obtain the corresponding first-stage unit commitment solutions.

(2) Fix the unit commitment solutions obtained in (1) and solve the second-stage problem repeatedly for 50 randomly generated instances to obtain the total cost for each approach.

To compare the performances between these two approaches, we introduce a penalty cost at the rate of $\$5,000$/MWh [7], for any power imbalance or transmission capacity/ramp-rate limit violation.

We first test the performances of the SO and SR approaches under various Budget% settings. We set the scenario size to be 5 and report the computational results in Table 3-2. The Budget% scenarios are given in the first column. The total costs (T.C.) obtained by each approach are reported in the third column. The UC costs (UC.C.) for each approach are given in the fourth column and the numbers of start-ups are given in the fifth column. Finally, the CPU times are recorded in the sixth column.

From Table 3-2, we have the following observations:

(1) First, we observe that when Budget% ≤ 5, the unit commitment decisions obtained from the SO and SR approaches are the same. This is because the uncertainty set is so small that the unit commitment decisions obtained from the SO approach are robust enough to accommodate the uncertainty, and there is no need to generate feasibility cuts in the SR approach. However, when Budget% ≥ 10, the UC decisions obtained from the SO approach are not feasible to some simulated load scenarios. But the UC decisions obtained from the SR approach are always feasible under different budget levels. Therefore, due to the penalty cost for the violation in the power balance or transmission capacity constraints, the SR approach incurs a smaller total cost than the SO approach when Budget% $\geq 10\%$. Moreover, when Budget% increases, the total cost gap between these two approaches increases. This result verifies that the proposed approach can provide a more robust solution as compared to the SO approach, especially when the system has more uncertainties.

Table 3-2. Comparison between SO and SR approaches

Budget%	Model	T.C.($)	UC.C.($)	# of Start-ups	Time(sec)
3%	SO	737,799	49,500	10	62
	SR	737,799	49,500	10	50
5%	SO	738,112	49,500	10	59
	SR	738,112	49,500	10	47
10%	SO	740,878	49,500	10	62
	SR	739,911	52,500	11	126
15%	SO	752,575	49,500	10	63
	SR	741,170	51,000	11	167
20%	SO	782,361	49,500	10	63
	SR	742,866	54,000	12	222

(2) Second, we observe that, as compared to the SO approach, the SR approach requires more generators committed to provide sufficient generation capacity to guarantee the supply meeting the load. As a result, the SR approach has a larger UC cost than the SO approach. This result also verifies that the proposed SR approach can provide a more robust solution as compared to the SO approach.

Next, we test the system performances of the SO and SR approaches under various scenario-size settings. We set Budget$\% = 20\%$ and test the scenario size $N = 1, 5, 10$, and 20, respectively. First, we observe that for the SO approach, the number of start-ups increases when the number of scenarios increases (e.g., the number of start-ups is 10 when $N = 1$ and this number increases to 11 when $N = 20$). On the other hand, in our proposed SR approach, the number of start-ups remains the same. Therefore, the proposed SR approach is more robust than the traditional SO approach. Second, we observe that as the scenario size increases, the total cost of the SO approach and the total cost gap between the SO and SR approaches decrease (e.g., the total cost of SO is $\$782,361$ and the total cost of SR is $\$742,866$ when $N = 1$; the total cost of SO is $\$758,496$ and the total cost of SR is $\$742,866$ when $N = 20$). This is because as the scenario size increases, the first-stage unit commitment solution obtained by the SO approach becomes more robust so that the system incurs less penalty cost

in the second stage. However, there is still a big gap between the two when $N = 20$, which verifies that it is necessary to add the robust uncertainty set to ensure system robustness of the day-ahead unit commitment decision.

3.5.3 Proposed Approach vs Robust Optimization Approach

We also compare the performances of our proposed SR approach with the traditional two-stage robust optimization (RO) approach under various Budget% settings. We set the scenario size for the SR approach to be 5 and summarize the computational results in Table 3-3. The Budget% settings, the total costs, the UC costs, the numbers of start-ups, and the CPU times are reported in the first, third, fourth, fifth and sixth columns respectively.

First, from our computational results, we observe that there are no penalty costs incurred for both the SR and RO approaches, which means that the unit commitment decisions for each approach are feasible for all generated scenarios. Second, from Table 3-3 we observe that our proposed SR approach commits less number of units in the first stage than the RO approach. That is, as compared to the RO approach, SR leads to smaller unit commitment and total costs. This result indicates that our proposed SR approach can generate a less conservative solution as compared to the RO approach while maintaining system robustness. Finally, we observe that, the CPU times for the RO approach are larger than our proposed SR approach for each tested instance, because the initial solution for the RO approach is worse than that of our proposed SR approach, and more optimality cuts are generated to make the algorithm converge.

3.6 Summary

In this chapter, we developed a unified stochastic and robust unit commitment model for ISOs/RTOs to perform reliability unit commitment runs, so as to achieve a robust and cost-effective unit commitment solution. Our proposed approach takes the advantages of both the stochastic and robust optimization approaches. The uncertainty set provided by the robust optimization approach ensures the robustness of the unit

Table 3-3. Comparison between RO and SR approaches

Budget%	Model	T.C.($)	UC.C.($)	# of Start-ups	Time (sec)
3%	RO	737,294	49,500	10	375
	SR	737,275	49,500	10	49
5%	RO	738,514	51,000	11	292
	SR	737,275	49,500	10	48
10%	RO	739,515	54,000	12	375
	SR	738,190	52,500	11	127
15%	RO	739,868	54,000	12	339
	SR	738,506	51,000	11	168
20%	RO	749,064	63,300	15	303
	SR	739,320	54,000	12	223

commitment decision. Meanwhile, the expected total cost in the objective function provides system operators flexibility to adjust the cost-effectiveness of the proposed approach. The proposed model can be solved efficiently by our proposed Benders' decomposition framework, which includes both feasibility and optimality cuts. Finally, computational experiments show the effectiveness of our proposed approach.

CHAPTER 4
DATA-DRIVEN RISK-AVERSE TWO-STAGE STOCHASTIC PROGRAM

4.1 Problem Description and Literature Review

Stochastic programming can effectively help decision making in an uncertain environment. Particularly, a specific class of stochastic programming has been studied extensively, which is called two-stage stochastic program. Typically, a two-stage stochastic program (denoted as SP) can be formulated as follows:

$$\begin{aligned} & \min_x c^T x + E[Q(x,\xi)] \\ \text{(SP)} \quad & s.t.\ x \in X, \end{aligned} \tag{4–1}$$

where x is the first-stage decision variable, in a convex set X, and $Q(x,\xi) = \min_{y(\xi)\in Y}$ $\{H(y(\xi)) : G(x,y(\xi)) \leq 0, \xi \in \Omega\}$ representing the second-stage problem in which the random variable ξ is on a probability space $(\Omega, \sigma(\Omega), \mathbb{P})$, where Ω is the sample space for ξ, $\sigma(\Omega)$ is the σ-algebra of Ω, and $\mathbb{P}$ is the associate probability distribution. In this formulation, the probability distribution is assumed known and significant research progress has been made in the theoretical analysis and developing efficient algorithms [9]. However, in practice, due to limited available information on the random parameters, it is generally difficult to precisely estimate the probability distribution $\mathbb{P}$. To address this issue, distributionally robust stochastic program has been proposed. In this approach, by defining an uncertainty set $\mathcal{D}$ of the true probability distribution $\mathbb{P}$, a distributionally robust two-stage stochastic program can be reformulated as follows (denoted as RA-SP), with the objective of minimizing the expected cost under the worst-case distribution in $\mathcal{D}$:

$$\begin{aligned} & \min_x c^T x + \max_{\mathbb{P}\in\mathcal{D}} E_{\mathbb{P}}[Q(x,\xi)] \\ \text{(RA-SP)} \quad & s.t.\ x \in X. \end{aligned} \tag{4–2}$$

One important factor for solving the distributionally robust stochastic program is the construction of $\mathcal{D}$. Several approaches have been proposed in the literature. For instance, in [25], the conic sets for the mean vector of observed data are considered; in [58], by restricting the density function f with l_∞ norm, the distributionally robust stochastic program can be reformulated as a CVaR problem and solved by a sample average approximation approach.

More recently, the data driven approach has been applied to construct the confidence set $\mathcal{D}$. That is, instead of knowing the true probability distribution $\mathbb{P}$, a series of historical data that are taken from $\mathbb{P}$ can be observed. With a reference distribution $\mathbb{P}_0$ determined by the historical data and a predefined metric $d(\mathbb{P}_0, \mathbb{P})$ to measure the distance between $\mathbb{P}_0$ and $\mathbb{P}$, the confidence set $\mathcal{D}$ can be represented as

$$\mathcal{D} = \{\mathbb{P} : d(\mathbb{P}, \mathbb{P}_0) \leq \theta\}, \tag{4–3}$$

where the tolerance θ is dependent on the number of historical data observed. Intuitively, $\mathcal{D}$ gets tighter around the true distribution $\mathbb{P}$ with more historical data observed. That is, θ becomes smaller, and accordingly RA-SP becomes less conservative. There has been research progress made by using the data driven approach to construct the distribution confidence uncertainty set $\mathcal{D}$. For instance, in [18], $\mathcal{D}$ is constructed based on the mean and covariance matrices of the random variables learned from the historical data. In addition, $\mathcal{D}$ can be constructed differently by defining different metrics. For example, ϕ-divergences, which are defined in the form $d_\phi(\mathbb{P}|\mathbb{P}_0) := \int_\Omega \phi(\frac{f(\xi)}{f_0(\xi)}) f_0(\xi)\, d\xi$ with reference density function f_0 corresponding to $\mathbb{P}_0$ and true density function f corresponding to $\mathbb{P}$, are usually used for measuring the distance between the reference distribution and the true distribution. In [6], ϕ-divergences are used for the metric d and the single-stage robust counterpart problem can be reformulated as a tractable problem. This approach has been extended to the two-stage stochastic program case in [38] and the chance constrained case in [29]. Besides, it is also well-known that

ϕ-divergences, especially KL-divergence $d_{KL}(\mathbb{P}|\mathbb{P}_0) := \int_\Omega \log(\frac{f(\xi)}{f_0(\xi)}) f_0(\xi) d\xi$, has been commonly applied in both machine learning and information theory, e.g., [46] and [62]. However, ϕ-divergences are in general not metrics since most ϕ-divergences do not satisfy the triangle inequality and even the symmetric property $d_\phi(\mathbb{P}|\mathbb{P}_0) = d_\phi(\mathbb{P}_0|\mathbb{P})$. Also, as indicated in [21] and the metrics' relationships in [24], there are no convergence results for general ϕ-divergences with empirical probability distribution as $\mathbb{P}_0$, for the case in which the true density function f is continuous.

In this chapter, we introduce a new class of metrics, which is called ζ-structure probability metrics class. Given two probability distributions $\mathbb{P}$ and $\mathbb{Q}$, the ζ-structure probability metrics are defined as follows:

$$d_\zeta(\mathbb{P}, \mathbb{Q}) = \sup_{h \in \mathcal{H}} \left| \int_\Omega h d\mathbb{P} - \int_\Omega h d\mathbb{Q} \right|, \tag{4–4}$$

where $\mathcal{H}$ is a class of real-valued bounded measurable functions on Ω. ζ-structure probability metrics class is first introduced in [79], and it has been applied in information theory [26], mathematical statistics [57], mass transportation problems [52], and several areas in computer science, including probabilistic concurrency, image retrieval, data mining, and bioinformatics [20]. The members in ζ-structure probability metrics class are Kantorovich metric, Fortet-Mourier metric, Total Variation metric, Bounded Lipschitz metric, and Kolmogorov/Uniform metric. We will introduce these metrics in details in Section 4.2. In our chapter, we utilize ζ-structure probability metrics to help construct the confidence uncertainty set for the probability distribution in RA-SP (i.e., Problem (4–2)). Then, we develop algorithms to solve RA-SP for both discrete and continuous true distribution cases. Finally, we show that RA-SP converges to SP (i.e., Problem (4–1)) as the number of historical data increases to infinity and further explore the value of data by investigating the convergence rate and showing numerical experiment results. Our contributions can be summarized as follows:

1. We apply a new class of metrics, ζ-structure probability metrics, to construct the confidence set $\mathcal{D}$ for the ambiguous distributions and explore the relationships among the members in the ζ-structure probability metrics class.

2. We provide frameworks to solve RA-SP for both discrete and continuous distribution cases. For the case in which the true distribution $\mathbb{P}$ is discrete, we can reformulate RA-SP as a traditional two-stage robust optimization problem. For the case in which $\mathbb{P}$ is continuous, we propose a sampling approach and obtain the lower and upper bounds of the optimal objective value of RA-SP. We further prove that both lower and upper bounds converge to the optimal objective value of RA-SP as the sample size goes to infinity.

3. We explore the convergence rates for both outer and inner loops. That is, for the outer loop, we prove that as the number of historical data goes to infinity, the risk-averse problem RA-SP converges to SP exponentially fast. In addition, for the inner loop for which the true distribution is continuous, as the sample size goes to infinity, we prove that both lower and upper bounds converge to the optimal objective value of RA-SP exponentially fast.

The remaining part of this chapter is organized as follows: In Section 4.2, we introduce the ζ-structure probability metrics class and study the relationships among the members in this class. In Section 4.3, we describe the solution approaches to solve RA-SP for both discrete and continuous true distribution cases, and explore the convergence rate for each case. In Section 4.4, we perform numerical studies on data-driven risk-averse newsvendor and facility location problems. Finally, in Section 4.5, we conclude our research.

4.2 ζ-Structure Probability Metrics

In this section, we introduce a new class of metrics: ζ-structure probability metrics. We first introduce the definitions of the metric members of that class, e.g., Kantorovich metric, Fortet-Mourier metric, Total Variation metric, Bounded Lipschitz metric, and Kolmogorov/Uniform metric. Then, we investigate the relationships among these metrics. We show that if the supporting space Ω is bounded, the Total Variation metric is a dominating metric in the class, i.e., the convergence with the Total Variation metric can guarantee the convergence with other metrics in the class.

4.2.1 Definition

As described in Section 4.1, for any two probability distributions $\mathbb{P}$ and $\mathbb{Q}$, the ζ-structure probability metrics are defined by $d_\zeta(\mathbb{P}, \mathbb{Q}) = \sup_{h\in\mathcal{H}} |\int_\Omega h d\mathbb{P} - \int_\Omega h d\mathbb{Q}|$, where $\mathcal{H}$ is a class of real-valued bounded measurable functions on Ω. In general, the ζ-structure metrics satisfy the properties of metrics, i.e., $d_\zeta(\mathbb{P}, \mathbb{Q}) = 0$ if and only if $\mathbb{P} = \mathbb{Q}$, $d_\zeta(\mathbb{P}, \mathbb{Q}) = d_\zeta(\mathbb{Q}, \mathbb{P})$ (symmetric property) and $d_\zeta(\mathbb{P}, \mathbb{Q}) \leq d_\zeta(\mathbb{P}, \mathbb{O}) + d_\zeta(\mathbb{O}, \mathbb{Q})$ for any probability distribution $\mathbb{O}$ (triangle inequality). In the following, we define $\rho(x, y)$ as the distance between random variables x and y and define d as the dimension of Ω. If random variable x follows distribution $\mathbb{P}$, we denote it as $\mathbb{P} = \mathcal{L}(x)$. Then, we derive different types of metrics based on different definitions of $\mathcal{H}$.

- **Kantorovich metric**: For Kantorovich metric (denoted as $d_K(\mathbb{P}, \mathbb{Q})$), $\mathcal{H} = \{h : \|h\|_L \leq 1\}$, where $\|h\|_L := \sup\{\frac{h(x)-h(y)}{\rho(x,y)} : x \neq y \text{ in } \Omega\}$. Many metrics known in statistics, measure theory, ergodic theory, and functional analysis, are special cases of the Kantorovich metric [67]. Kantorovich metric also has many applications in transportation theory [52], and in computer science including probabilistic concurrency, image retrieval, data mining, and bioinformatics [20].

- **Fortet-Mourier metric**: For Fortet-Mourier metric (denoted as $d_{FM}(\mathbb{P}, \mathbb{Q})$), $\mathcal{H} = \{h : \|h\|_C \leq 1\}$, where $\|h\|_C := \sup\{\frac{h(x)-h(y)}{c(x,y)} : x \neq y \text{ in } \Omega\}$ and $c(x, y) = \rho(x, y)\max\{1, \rho(x, a)^{p-1}, \rho(y, a)^{p-1}\}$ for some $p \geq 1$ and $a \in \Omega$. Note here that when $p = 1$, Fortet-Mourier metric is the same as Kantorovich metric. Therefore, Fortet-mourier metric is usually used as a generalization of Kantorovich metric, with the applications on mass transportation problems [52].

- **Total Variation metric**: For Total Variation metric (denoted as $d_{TV}(\mathbb{P}, \mathbb{Q})$), $\mathcal{H} = \{h : \|h\|_\infty \leq 1\}$, where $\|h\|_\infty := \sup_{x\in\Omega} |h(x)|$. Another definition of the Total Variation metric is $d_{TV}(\mathbb{P}, \mathbb{Q}) := 2\sup_{B\in\sigma(\Omega)} |\mathbb{P}(B) - \mathbb{Q}(B)|$. The total variation metric has a coupling characterization (detailed proofs are shown in [37]):

$$d_{TV}(\mathbb{P}, \mathbb{Q}) = 2\inf\{\Pr(X \neq Y) : \mathbb{P} = \mathcal{L}(X), \mathbb{Q} = \mathcal{L}(Y)\}. \tag{4–5}$$

  The total variance metric can be applied in information theory [16] and in studying the ergodicity of Markov Chains [45]. Later on, we will prove that the convergence with respect to the Total Variation metric implies the convergence with respect to other metrics in the general ζ-structure probability metrics class when Ω is bounded.

- **Bounded Lipschitz metric**: For Bounded Lipschitz metric (denoted as $d_{BL}(\mathbb{P}, \mathbb{Q})$), $\mathcal{H} = \{h : \|h\|_{BL} \leq 1\}$, where $\|h\|_{BL} := \|h\|_L + \|h\|_\infty$. Bounded Lipschitz metric is usually used to prove the convergence of probability measures in the weak topology [22].

- **Uniform/Kolmogorov metric**: For Uniform metric (also called Kolmogorov metric, denoted as $d_U(\mathbb{P}, \mathbb{Q})$), $\mathcal{H} = \{I_{(-\infty,t]}, t \in R^d\}$. The Uniform metric is often used in proving the classical central limit theorem, and commonly utilized in the Kolmogorov-Smirnov statistic for hypothesis testing [56]. According to the definition, we have $d_U(\mathbb{P}, \mathbb{Q}) = \sup_t |\mathbb{P}(x \leq t) - \mathbb{Q}(x \leq t)|$.

- **Wasserstein metric**: Wasserstein metric is defined as $d_W(\mathbb{P}, \mathbb{Q}) := \inf_\pi\{E_\pi[\rho(X, Y)] : \mathbb{P} = \mathcal{L}(X), \mathbb{Q} = \mathcal{L}(Y)\}$, where the infimum is taken over all joint distributions π with marginals $\mathbb{P}$ and $\mathbb{Q}$. Although Wasserstein metric is not a member in the general ζ-structure probability metrics class, by the Kantorovich-Rubinstein theorem [33], the Kantorovich metric is equivalent to the Wasserstein metric. In particular, when $\Omega = R$,

$$d_W(\mathbb{P}, \mathbb{Q}) = \int_{-\infty}^{+\infty} |F(x) - G(x)| dx, \tag{4–6}$$

  where $F(x)$ and $G(x)$ are the distribution functions derived from $\mathbb{P}$ and $\mathbb{Q}$ respectively. This conclusion holds following the argument that $\inf_\pi\{E_\pi[\rho(X, Y)] : \mathbb{P} = \mathcal{L}(X), \mathbb{Q} = \mathcal{L}(Y)\} = \int_0^1 |F^{-1}(t) - G^{-1}(t)| dt$, as stated in Theorem 6.0.2 in [3] and $\int_0^1 |F^{-1}(t) - G^{-1}(t)| dt = \int_{-\infty}^{+\infty} |F(x) - G(x)| dx$. Wasserstein metric also has wide applications on transportation problems [52].

4.2.2 Relationships among Metrics

In this subsection, we explore the relationships among the members of ζ-structure probability metrics. Based on the relationships, we demonstrate that the Total Variation metric is the dominating metric if the supporting space Ω is bounded. By proving that the Total Variation metric is equal to the $\mathcal{L}_1$ metric when the probability density exists, we can also claim that the $\mathcal{L}_1$ metric is the dominating metric of the ζ-structure probability metrics class.

First, we explore the relationships between the Total Variation metric and the Wasserstein metric. We denote B as the diameter of Ω, and we have the following lemma.

Lemma 1. *The relationships between the Total Variation metric and the Wasserstein metric are as follows:*

- *If* Ω *is bounded,* $2d_W(\mathbb{P}, \mathbb{Q}) \leq B \cdot d_{TV}(\mathbb{P}, \mathbb{Q})$*;*
- *If* Ω *is a finite set,* $d_{min} \cdot d_{TV}(\mathbb{P}, \mathbb{Q}) \leq 2d_W(\mathbb{P}, \mathbb{Q})$*, where* $d_{min} = \min_{x \neq y} \rho(x, y)$*.*

Proof. See the detailed proof in Theorem 4 in [24]. □

Next, we study the relationships among the Bounded Lipschitz metric, the Wasserstein metric and the Total Variation metric. Since $\|h\|_{BL} := \|h\|_L + \|h\|_\infty$, the feasible region of h for the Bounded Lipschitz metric is more restrictive than the one for the Kantorovich metric and the one for the Total Variation metric. Therefore we have the following lemma:

Lemma 2. *The relationships among the Bounded Lipschitz, Wasserstein, and Total Variation metrics are as follows:*

- $d_{BL}(\mathbb{P}, \mathbb{Q}) \leq d_W(\mathbb{P}, \mathbb{Q})$*,*
- $d_{BL}(\mathbb{P}, \mathbb{Q}) \leq d_{TV}(\mathbb{P}, \mathbb{Q})$*.*

Moreover, the relationship between the Total Variation metric and the Uniform metric can be obtained by the definitions of each metric. The Total Variation metric is to find a set B among all the Borel set of $\sigma(\Omega)$, to minimize the value of $2|\mathbb{P}(B) - \mathbb{Q}(B)|$. But the Uniform metric is to find a set B among all the sets in the form $(-\infty, x]$, to minimize the value of $|\mathbb{P}(B) - \mathbb{Q}(B)|$. Since set $(-\infty, x]$ is a borel set, we have:

Lemma 3. *The relationship between the Total Variation metric and the Uniform metric is:* $2d_U(\mathbb{P}, \mathbb{Q}) \leq d_{TV}(\mathbb{P}, \mathbb{Q})$*.*

Finally, we explore the relationships between the Wasserstein metric and the Fortet-Mourier metric. For notation brevity, we let $L = \max\{1, B^{p-1}\}$. We have the following lemma:

Lemma 4. *The relationships between the Wasserstein metric and the Fortet-Mourier metric are as follows:*

- $d_W(\mathbb{P}, \mathbb{Q}) \leq d_{FM}(\mathbb{P}, \mathbb{Q})$,
- $d_{FM}(\mathbb{P}, \mathbb{Q}) \leq L d_W(\mathbb{P}, \mathbb{Q})$.

Proof. The first inequality is obvious by following the definitions of the Wasserstein metric and the Fortet-Mourier metric. For the second inequality, for the Fortet-Mourier metric, we have $|h(x) - h(y)| \leq c(x, y) = \rho(x, y) \max\{1, \rho(x, a)^{p-1}, \rho(y, a)^{p-1}\} \leq \rho(x, y)L$, where the equation holds following the definition of $c(x, y)$. Now we can observe

$$\begin{aligned} d_{FM}(\mathbb{P}, \mathbb{Q}) \leq & \sup_{h:|h(x)-h(y)|\leq L\cdot\rho(x,y)} |\int_\Omega h d\mathbb{P} - \int_\Omega h d\mathbb{Q}| \\ &= L \cdot \sup_{h:|h(x)-h(y)|\leq L\cdot\rho(x,y)} |\int_\Omega \frac{h}{L} d\mathbb{P} - \int_\Omega \frac{h}{L} d\mathbb{Q}| \\ &= L \cdot \sup_{g:|g(x)-g(y)|\leq \rho(x,y)} |\int_\Omega g d\mathbb{P} - \int_\Omega g d\mathbb{Q}| \\ &= L \cdot d_W(\mathbb{P}, \mathbb{Q}). \end{aligned}$$

Then the second inequality holds. □

To summarize, the relationships among the members of ζ-structure probability metrics class are shown in Figure 4-1.

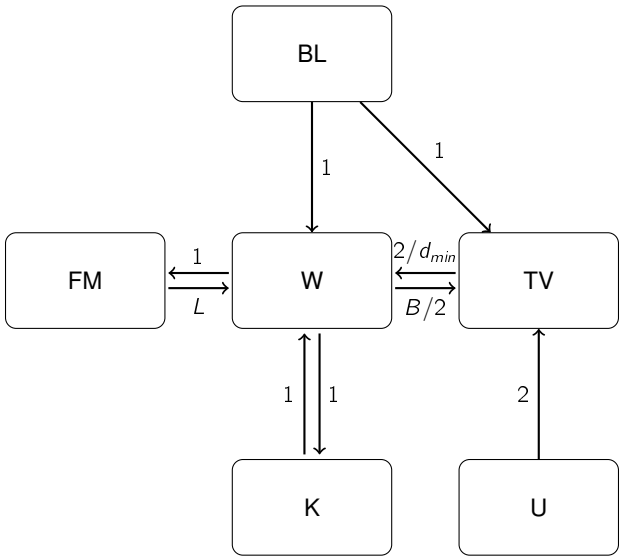

Figure 4-1. Relationships among members of ζ-structure probability metrics class

Based on Lemmas 1 to 4, we conclude the following proposition without proof.

Proposition 4.1. *If the support Ω is bounded, the Kantorovich metric, Fortet-Mourier metric, Bounded Lipschitz metric and Kolmogorov metric are dominated by the Total Variation metric.*

Moreover, if the density functions corresponding to $\mathbb{P}$ and $\mathbb{Q}$ exist (denote as $f(x)$ and $g(x)$), then we can define the $\mathcal{L}_1$ metric of the densities f, g as $d_{\mathcal{L}_1}(f, g) := \int_{R^d} |f(x) - g(x)| dx$. In the following lemma, we show the relationship between the $\mathcal{L}_1$ metric and the Total Variation metric.

Lemma 5. $d_{\mathcal{L}_1}(f, g) = d_{TV}(\mathbb{P}, \mathbb{Q})$.

Proof. First, letting $A := \{x \in \Omega : f(x) > g(x)\}$, we have

$$\begin{aligned}\int_{R^d} |f(x) - g(x)| dx &= \int_A |f(x) - g(x)| dx + \int_{A^c} |f(x) - g(x)| dx \\ &= \int_A (f(x) - g(x)) dx + \int_{A^c} (g(x) - f(x)) dx \\ &= 2 \int_A (f(x) - g(x)) dx = 2(\mathbb{P}(A) - \mathbb{Q}(A)).\end{aligned}$$

The last step is because of $\int_{R^d} (f(x) - g(x)) dx = 0$. Second, for any $B \in \sigma(\Omega)$,

$$\begin{aligned}|\mathbb{P}(B) - \mathbb{Q}(B)| &= |\int_B f(x) dx - \int_B g(x) dx| \\ &= |\int_{B \cap A} (f(x) - g(x)) dx + \int_{B \cap A^c} (f(x) - g(x)) dx| \\ &\leq \max\{\int_{B \cap A} (f(x) - g(x)) dx, \int_{B \cap A^c} (g(x) - f(x)) dx\} \\ &\leq \max\{\int_A (f(x) - g(x)) dx, \int_{A^c} (g(x) - f(x)) dx\} \\ &= \frac{1}{2} \int |f(x) - g(x)| dx.\end{aligned}$$

Therefore, we have $d_{TV}(\mathbb{P}, \mathbb{Q}) = d_{\mathcal{L}_1}(f, g)$. □

With the above lemmas and proposition, we explore the methodologies to solve RA-SP (i.e., Problem (4–2)) and study the convergence rates in the next section.

4.3 Solution Methodology

In this section, we investigate the solution methodologies to solve the data-driven risk-averse two-stage stochastic program (i.e., RA-SP as described in (4–2)). For theoretical convenience, we assume the supporting space Ω is bounded, which is not a restrictive assumption in practice. The section is organized by answering the following important questions:

1. How to determine the reference distribution $\mathbb{P}_0$?
2. How to represent the value of θ depending on the amount of historical data, i.e., the convergence rate?
3. How to solve the problem with respect to different ζ-structure probability metrics?

We answer these questions by considering two cases: (i) the true probability distribution is discrete and (ii) the true probability distribution is continuous.

4.3.1 Discrete Case

For the case in which the true distribution is discrete, we first consider the case that the supporting space Ω is finite, e.g., $\{\xi^1, \xi^2, \cdots, \xi^N\}$. For the case that the supporting space Ω is infinite, we can use the framework of continuous probability distribution case, which we will discuss later. Given M historical data $\xi_0^1, \xi_0^2, \cdots, \xi_0^M$, to estimate the reference distribution $\mathbb{P}_0$, we consider the empirical distribution of historical data, i.e., the cumulative distribution function (CDF) that puts mass $1/M$ at each data point $\xi_0^i, i = 1 \cdots, M$. Formally, the empirical distribution is defined as

$$\mathbb{P}_0(x) = \frac{1}{M} \sum_{i=1}^{M} \delta_{\xi_0^i}(x), \tag{4–7}$$

where $\delta_{\xi_0^i}(x)$ is an indicator variable equal to 1 if $\xi_0^i \leq x$ and 0 otherwise. In this case, since the supporting space is discrete, the reference distribution $\mathbb{P}_0$ can be represented by its mass probability $p_0^1, p_0^2, \cdots, p_0^N$, where p_0^i is equal to the number of historical data matching ξ^i divided by M.

After identifying the reference distribution $\mathbb{P}_0$, we discuss the value of θ, i.e., the convergence rate of the empirical distribution to the true distribution of the discrete case. We first study the convergence rate of the Total Variation metric for the discrete distribution, and then we explore the convergence rates of other metrics in the class for a more general case, i.e., the true distribution is not limited to be discrete (these results will be used for later on continuous case convergence rate analysis).

For the Total Variation metric, if the true distribution is discrete, Pinsker's inequality [17] shows $d_{TV}(\mathbb{P}_0, \mathbb{P}) \leq \sqrt{d_{KL}(\mathbb{P}|\mathbb{P}_0)}$, where $d_{KL}(\mathbb{P}|\mathbb{P}_0)$ is the discrete case KL-divergence defined as $\sum_i \ln(\frac{p_i^0}{p_i})p_i^0$. Since [50] shows that $d_{KL}(\mathbb{P}|\mathbb{P}_0)$ converges in measure to a χ-square distributed random variable as number of historical data M goes to infinity, we can claim that the convergence rate of $d_{TV}(\mathbb{P}_0, \mathbb{P})$ is bounded by that of a χ distributed random variable as M goes to infinity. Meanwhile, note here that for the case in which the true distribution is continuous, the total variation metric does not converge as described in [21], which is also the reason that the KL divergence does not converge as indicated in Section 1.

Next, we explore the convergence rate of the Kolmogorov metric for the general distribution case (either discrete or continuous). The convergence property and the convergence rate can be obtained by utilizing the following Dvoretzky-Kiefer-Wolfowitz inequality.

Proposition 4.2 (Dvoretzky-Kiefer-Wolfowitz)**.** *For the $d = 1$ case,* $\Pr(d_U(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - 2e^{-2M\epsilon^2}$*. For the $d > 1$ case, for any $\alpha > 0$, there exists a constant number C_α such that* $\Pr(d_U(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - C_\alpha e^{-(2-\alpha)M\epsilon^2}$*.*

Remark: A tighter rate has been obtained in [1] for the $d > 1$ case in which $\Pr(d_U(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - C\epsilon^{2(d-1)}M^{d-1}e^{-2M\epsilon^2}$, for some $C \geq 0$.

Finally, we explore the convergence rate of the Wasserstein metric (note here the rate also applies to the Kantorovich metric) for the general distribution case, and

therefore we can obtain the convergence rates of the Fortet-Mourier and Bounded Lipschitz metrics accordingly.

Proposition 4.3. *If* $d = 1$*, then* $\Pr(d_W(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - 2e^{-2M\epsilon^2}$.

Proof. For $d = 1$, based on the definition of the Wasserstein metric, $d_W(\mathbb{P}_0, \mathbb{P}) = \int_{-\infty}^{+\infty} |F(x) - G(x)|dx$, where F and G are distribution functions derived from $\mathbb{P}_0$ and $\mathbb{P}$ respectively. According to the Hölder's inequality $\|fg\|_1 \leq \|f\|_p \|g\|_q$ with $1/p + 1/q = 1$, letting $f = F(x) - G(x)$, $g = 1$, $p = \infty$, and $q = 1$, we have

$$\int_{-\infty}^{+\infty} |F(x) - G(x)|dx \leq \sup_x |F(x) - G(x)|.$$

Moreover, according to the Dvoretzky-Kiefer-Wolfowitz inequality described in Proposition 4.2, we have $\Pr(\sup_x |F(x) - G(x)| \leq \epsilon) \geq 1 - 2e^{-2M\epsilon^2}$. Therefore, $\Pr(d_W(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - 2e^{-2M\epsilon^2}$. □

We then can immediately obtain the convergence rates for the Fortet-Mourier and Bounded Lipschitz metrics for the $d = 1$ case, following the relationships among the ζ-structure probability metrics as described in Subsection 4.2.2.

Corollary 1. *If* $d = 1$*, then* $\Pr(d_{FM}(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - 2e^{-\frac{2M}{L^2}\epsilon^2}$ *and* $\Pr(d_{BL}(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - 2e^{-2M\epsilon^2}$.

Next, we prove that for a general dimension d, $\mathbb{P}_0$ also converges to $\mathbb{P}$ exponentially fast with respect to the Wasserstein metric. In [10], it has shown for $\Omega = R^K$, the convergence rate of the empirical distribution to the true distribution is

$$P(d_W(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - \exp(-\frac{\lambda}{2}M\epsilon^2),$$

for some $\lambda \geq 0$. Motivated by this, we now explore an exact convergence rate for a bounded space Ω. It is shown in "Particular case 5" in [11] that

$$d_W(\mu, \mathbb{P}_0) \leq diam(\Omega)\sqrt{2d_{KL}(\mu|\mathbb{P}_0)} \tag{4–8}$$

holds for $\forall \mu \in \mathbf{P}(\Omega)$, where $\mathbf{P}(\Omega)$ is the set of all probability measures defined on Ω. With inequality (4–8), we have the following proposition.

Proposition 4.4. *For a general $d \geq 1$, we have*

$$P(d_W(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - \exp(-\frac{\epsilon^2}{2B^2}M).$$

Proof. Let set

$$\mathcal{B} := \{\mu \in \mathbf{P}(\Omega) : d_W(\mu, \mathbb{P}) \geq \epsilon\} \tag{4–9}$$

and $\mathbf{C}(\Omega)$ be the collection of all bounded continuous functions ϕ: $\Omega \to R$, with the supremum norm, i.e., $\|\phi\|_\infty = \sup_x |\phi(x)|$. Therefore, we have

$$\begin{aligned}
\Pr(d_W(\mathbb{P}, \mathbb{P}_0) \geq \epsilon) &= Pr(\mathbb{P}_0 \in \mathcal{B}) \\
&\leq Pr\left(\int_\Omega \phi d\mathbb{P}_0 \geq \inf_{\mu\in\mathcal{B}} \int_\Omega \phi d\mu\right) \\
&\leq \exp\left(-M \inf_{\mu\in\mathcal{B}} \int_\Omega \phi d\mu\right) E\left(e^{M\int_\Omega \phi d\mathbb{P}_0}\right) \qquad (4–10) \\
&= \exp\left(-M \inf_{\mu\in\mathcal{B}} \left\{\int_\Omega \phi d\mu - \frac{1}{M}\log E\left(e^{M\int_\Omega \phi d\mathbb{P}_0}\right)\right\}\right) \\
&= \exp\left(-M \inf_{\mu\in\mathcal{B}} \left\{\int_\Omega \phi d\mu - \frac{1}{M}\log E\left(e^{\sum_{i=1}^M \phi(\xi^i)}\right)\right\}\right) \qquad (4–11) \\
&= \exp\left(-M \inf_{\mu\in\mathcal{B}} \left\{\int_\Omega \phi d\mu - \log \int_\Omega e^\phi d\mathbb{P}_0\right\}\right), \qquad (4–12)
\end{aligned}$$

for $\forall \phi \in \mathbf{C}(\Omega)$, where inequality (4–10) follows from the Chebyshev's exponential inequality, equation (4–11) follows from the definition of $\mathbb{P}_0$, and equation (4–12) follows from the assumption that the historical data are i.i.d. drawn. Considering $\Delta(\mu) := \sup_{\phi\in\mathcal{C}(\Omega)} \int_\Omega \phi d\mu - \log \int_\Omega e^\phi d\mathbb{P}_0$, following the continuity and boundedness from the definition of $\mathbf{C}(\Omega)$, there exist a series ϕ_n such that $\lim_{n\to\infty} \int_\Omega \phi_n d\mu - \log \int_\Omega e^{\phi_n} d\mathbb{P}_0 = \Delta(\mu)$. Therefore, for any small positive number $\epsilon' > 0$, there exists an N_0 such that $\Delta(\mu) - \int_\Omega \phi_n d\mu - \log \int_\Omega e^{\phi_n} d\mathbb{P}_0 \leq \epsilon'$ for any $n \geq N_0$. Therefore, according to (4–12) by substituting ϕ with ϕ_n, we have

$$Pr(\mathbb{P}_0 \in \mathcal{B}) \leq \exp\left(-M \inf_{\mu\in\mathcal{B}} \left\{\int_\Omega \phi_n d\mu - \log \int_\Omega e^{\phi_n} d\mathbb{P}_0\right\}\right)$$

$$\leq \exp\left(-M \inf_{\mu\in\mathcal{B}}(\Delta(\mu)-\epsilon')\right).$$

According to Lemma 6.2.13 in [19], we know that $\Delta(\mu) = d_{KL}(\mu|\mathbb{P}_0)$. For the case $\mu \in \mathcal{B}$, following the definition of set $\mathcal{B}$ in (4–9), we have $d_W(\mu, \mathbb{P}_0) \geq \epsilon$. In addition, following (4–8) we have $d_{KL}(\mu|\mathbb{P}_0) \geq \epsilon^2/(2B^2)$. Therefore,

$$Pr(\mathbb{P}_0 \in \mathcal{B}) \leq \exp\left(-M\left(\frac{\epsilon^2}{2B^2}-\epsilon'\right)\right).$$

Let $\epsilon' = \delta/M$ for any arbitrary small positive δ, then

$$\Pr(d_W(\mathbb{P}, \mathbb{P}_0) \geq \epsilon) = Pr(\mathbb{P}_0 \in \mathcal{B}) \leq \exp\left(-\frac{\epsilon^2}{2B^2}M + \delta\right).$$

Since δ is arbitrarily small, we have $P(d_W(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - \exp(-\frac{\epsilon^2}{2B^2}M)$. □

Similarly, the convergence rate also applies for the Kantorovich metric, and moreover, we can obtain the convergence rates for the Fortet-Mourier and Bounded Lipschitz metrics for the general dimension case as follows.

Corollary 2. *For a general $d \geq 1$, we have $P(d_{FM}(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - \exp(-\frac{\epsilon^2}{2B^2L^2}M)$ and $\Pr(d_{BL}(\mathbb{P}_0, \mathbb{P}) \leq \epsilon) \geq 1 - \exp(-\frac{\epsilon^2}{2B^2}M)$.*

With the convergence rates, we can calculate the value of θ accordingly. Take the convergence rate for the Wasserstein metric obtained by Proposition 4.4 for example. Assuming the confidence level is set to be η, that is, $P(d_W(\mathbb{P}_0, \mathbb{P}) \leq \theta) \geq 1 - \exp(-\frac{\theta^2}{2B^2}M) = \eta$, then we can obtain $\theta = \sqrt{2B\log(\frac{1}{1-\eta})/M}$. Similarly, we can calculate the value of θ for different metrics based on Corollary 2.

Next, we explore the methodology to solve the problem. Assuming $\Omega = \{\xi^1, \xi^2, \cdots, \xi^N\}$, RA-SP (i.e., Problem (4–2)) can be reformulated as:

$$\begin{aligned} &\min_x\ c^T x + \max_{p_i} \sum_{i=1}^{N} p_i Q(x, \xi^i) \\ &s.t.\ \sum_i p_i = 1, \end{aligned} \tag{4–13}$$

$$\max_{h_i} \sum_{i=1}^{N} h_i p_i^0 - \sum_{i=1}^{N} h_i p_i \le \theta, \quad \forall h_i : \|h\|_\zeta \le 1. \tag{4–14}$$

For $\|h\|_\zeta \le 1$, accordingly, we have

- Kantorovich: $|h_i - h_j| \le \rho(\xi^i, \xi^j), \quad \forall i, j,$
- Fortet-Mourier: $|h_i - h_j| \le \rho(\xi^i, \xi^j) \max\{1, \rho(\xi^i, a)^{p-1}, \rho(\xi^j, a)^{p-1}\}, \quad \forall i, j,$
- Bounded-Lipschitz: $|h_i - h_j| \le \rho(\xi^i, \xi^j), |f_i| \le 1, \quad \forall i, j,$
- Total Variation: $|h_i| \le 1, \quad \forall i.$

For the above four metrics, the constraints can be summarized as $\sum_i a_{ij} h_i \le b_j, j = 1, \cdots, J$. First, we develop the reformulation of constraint (4–14). Considering the problem

$$\begin{aligned} \max_{h_i} \quad & \sum_{i=1}^{N} h_i p_i^0 - \sum_{i=1}^{N} h_i p_i \\ s.t. \quad & \sum_{i}^{N} a_{ij} h_i \le b_j, \quad j = 1, \cdots, J, \end{aligned}$$

we can get its dual formulation as follows:

$$\begin{aligned} \min \quad & \sum_{j=1}^{J} b_j u_j \\ s.t. \quad & \sum_{j=1}^{J} a_{ij} u_j \ge p_i^0 - p_i, \forall i = 1, \cdots, N, \end{aligned}$$

where u is the dual variable. Therefore, for the discrete distribution case, RA-SP can be reformulated as follows (denoted by FR-M):

$$\begin{aligned} \min_x \; & c^T x + \max_{p_i} \sum_{i=1}^{N} p_i G(x, \xi^i) \\ s.t. \; & \sum_{i=1}^{N} p_i = 1, \\ & \sum_{j=1}^{J} b_j u_j \le \theta, \end{aligned}$$

$$\sum_{j=1}^{J} a_{ij} u_j \geq p_i^0 - p_i, \forall i = 1, \cdots, N.$$

For the uniform metric, we can obtain the reformulation directly from the definition of Uniform metric (denoted by FR-U):

$$\begin{aligned} &\min\ c^T x + \max_{p_i} \sum_{i=1}^{N} p_i G(x, \xi^i) \\ &s.t. \sum_i p_i = 1, \\ &\quad |\sum_{i=1}^{j} p_i^0 - p_i| \leq \theta, \forall j = 1, \cdots, J. \end{aligned}$$

Next, we summarize the algorithm to solve the case in which the true distribution is discrete.

Algorithm 1 Algorithm for the discrete case

Input: Historical data $\xi^1, \xi^2, \cdots, \xi^N$ drawn i.i.d. from the true distribution and confidence level η.
Output: The objective value of the risk-averse problem RA-SP.
Step 1: Obtain the reference distribution $\mathbb{P}_0(x) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi^i}(x)$ and the value of θ based on the historical data.
Step 2: Use the reformulation FR-M or FR-U to solve the problem.
Step 3: Output the solution.

4.3.2 Continuous Case

In this section, we discuss the case in which the true distribution is continuous. Similarly, we develop the methodologies to solve the continuous case by answering the above three questions respectively. For the continuous case, we consider the reference probability density function (pdf) instead of the reference cumulative distribution function (cdf), and we use Kernel Density Estimation to construct the reference pdf. Assuming we observe n samples ξ^1,ξ^2,$\cdots$, ξ^n, which are drawn from the true distribution, the kernel density function is defined as

$$f_n(x) = \frac{1}{n h_n^d} \sum_{i=1}^{n} K(\frac{x - \xi^i}{h_n}), \tag{4–15}$$

where $K(x) = \prod_{k=1}^{d} k(x_k)$, in which k is a Borel measurable function (kernel) with dimension 1, satisfying $k \geq 0, \int_R k(x)dx = 1$. Note here that $K(x)$ also satisfies $K \geq 0, \int_{R^d} K(x)dx = 1$. By using the kernel density estimation, we can see that the reference probability distribution $\mathbb{P}_0$ derived from the density function $f_n(x)$ is a continuous distribution.

Next, we analyze the convergence properties and the convergence rates of the reference distribution to the true distribution, corresponding to different metrics in ζ-structure probability metrics class, by using kernel density estimation. Under mild assumptions, we can prove that the reference density distribution $\mathbb{P}_0$ derived from f_n converges to the true probability distribution exponentially fast, under different metrics.

Proposition 4.5. *If* $\lim_{n\to\infty} h_n = 0$ *and* $\lim_{n\to\infty} nh_n^{2d} = \infty$*, then for all* $\epsilon > 0$, $\Pr(|f_n(x) - E[f_n(x)]|_\infty > \epsilon) \leq 2\exp\{-nh_n^{2d}\epsilon^2/(2V^2)\}$ *for every* x*, where* $V = \sup_y K(y)$.

Proof. We let $Z_i = K(\frac{x-\xi^i}{h_n})/h_n^d$, then based on the definition of kernel density function (4–15), we have $E[f_n(x)] = E[\frac{1}{n}\sum_{i=1}^{n} Z_i] = E[Z_i]$ since that $\xi^i, i = 1, \cdots, N$ are drawn i.i.d.. Besides, it is easy to observe that

$$|Z_i| = |K(\frac{x-\xi^i}{h_n})/h_n^d| \leq \frac{V}{h_n^d}. \tag{4–16}$$

According to the Hoeffding's inequality, we have

$$\begin{aligned}\Pr(|f_n(x) - E[f_n(x)]| > \epsilon) &= \Pr(|\frac{1}{n}\sum_{i=1}^{n} Z_i - E[Z_i]| > \epsilon) \\ &\leq 2\exp\{-2n\epsilon^2/(\frac{2V}{h_n^d})^2\} \\ &= 2\exp\{-nh_n^{2d}\epsilon^2/(2V^2)\}.\end{aligned}$$

Thus we conclude the proof. □

Proposition 4.6. *If the true density function* $f(x)$ *is bounded and continuous almost everywhere, then* $\lim_{n\to\infty} |E[f_n(x)] - f(x)|_\infty = 0$.

Proof. On one hand, we have $E(f_n(x)) = \int_{R^d} h^{-d} K(\frac{x-y}{h}) f(y) dy = \int_{R^d} K(u) f(x-uh) du$. On the other hand, based on the definition of $K(u)$, we have $\int_{R^d} K(u) du = 1$, and therefore $f(x) = \int_{R^d} K(u) f(x) du$. Then,

$$|E[f_n(x)] - f(x)| = \left| \int_{R^d} K(u)(f(x-uh) - f(x)) du \right|$$
$$\leq \left| \int_{\|u\| \leq A} K(u)(f(x-uh) - f(x)) du \right| + \left| \int_{\|u\| \geq A} K(u)(f(x-uh) - f(x)) du \right|, \quad (4\text{–}17)$$

for any value $A \geq 0$. For the second item in (4–17), we can observe that

$$\left| \int_{\|u\| \geq A} K(u)(f(x-uh) - f(x)) du \right| \leq 2 \sup_{x \in R^d} f(x) \left| \int_{\|u\| \geq A} K(u) du \right|.$$

In addition, since $K(u) \geq 0$ and $\int_{u \in R^d} K(u) du = 1$, we can always choose A large enough to make $\sup_{x \in R^d} f(x) | \int_{\|u\| \geq A} K(u) du|$ less than any arbitrary ϵ. For the first item of (4–17), we have

$$\left| \int_{\|u\| \leq A} K(u)(f(x-uh) - f(x)) du \right| \leq \sup_{x \in R^d} |K(x)| \left| \int_{\|u\| \leq A} (f(x-uh) - f(x)) du \right|.$$

Since f is bounded and continuous almost everywhere, there exists a constant C such that $|f(x-uh) - f(x)| \leq C\|uh\|$ almost everywhere. Without considering the set with measure 0, we have $| \int_{\|u\| \leq A} K(u)(f(x-uh) - f(x)) du| \leq CAh \sup_{x \in R^d} |K(x)|$. Similarly, for a given A, $CAh \sup_{x \in R^d} |K(x)|$ can be less than any arbitrary ϵ when h is very small. So $|E[f_n(x)] - f(x)|_\infty \to 0$ when h goes to zero, so as n goes to infinity. □

With the above two propositions, we can estimate the convergence rate for the Total Variation metric as follows:

Theorem 4.1. *If the true density function $f(x)$ is bounded and continuous almost everywhere, $\lim_{n \to \infty} h_n = 0$, and $\lim_{n \to \infty} nh_n^{2d} = \infty$, then for all $\epsilon > 0$, there exists n_0 such that*

$$\Pr(d_{TV}(\mathbb{P}_0, \mathbb{P}) > \epsilon) \leq 2 \exp\{-nh_n^{2d} \epsilon^2 / (2V^2)\} \quad (4\text{–}18)$$

for $n \geq n_0$, where $V = \sup_y K(y)$ and $\mathbb{P}_0, \mathbb{P}$ are probability distributions derived from f_n and f respectively.

Proof. According to the Hölder's inequality, we have $d_{\mathcal{L}_1}(f_n, f) \leq \sup_x |f_n(x) - f(x)|$. According to Lemma 5, we have $d_{\mathcal{L}_1}(f_n, f) = d_{TV}(\mathbb{P}_0, \mathbb{P})$. Based on Propositions 4.5 and 4.6, we prove the theorem. □

With Theorem 4.1, it is easy to derive the convergence rates for other metrics based on Lemmas 1 to 4.

Corollary 3. *If the true density function $f(x)$ is bounded and continuous almost everywhere, $\lim_{n\to\infty} h_n = 0$, and $\lim_{n\to\infty} nh_n^d = \infty$, then for all $\epsilon > 0$, there exists an n_0 such that*

$$\Pr(d_W(\mathbb{P}_0, \mathbb{P}) > \epsilon) \leq 2\exp\{-nh_n^{2d}\epsilon^2/(2B^2V^2)\}, \tag{4–19}$$

$$\Pr(d_{BL}(\mathbb{P}_0, \mathbb{P}) > \epsilon) \leq 2\exp\{-nh_n^{2d}\epsilon^2/(2V^2)\}, \tag{4–20}$$

$$\Pr(d_{FM}(\mathbb{P}_0, \mathbb{P}) > \epsilon) \leq 2\exp\{-nh_n^{2d}\epsilon^2/(2(BLV)^2)\}, \tag{4–21}$$

$$\Pr(d_U(\mathbb{P}_0, \mathbb{P}) > \epsilon) \leq 2\exp\{-2nh_n^{2d}\epsilon^2/V^2\}, \tag{4–22}$$

for $n \geq n_0$, where $V = \sup_y K(y)$ and $\mathbb{P}_0, \mathbb{P}$ are probability distributions derived from f_n and f respectively.

Given that the reference distribution converges to the true distribution exponentially fast, next, we explore the convergence property of the optimal solution and the optimal objective value of RA-SP to the optimal solution and the optimal objective value of SP, i.e., Problem (4–1). We have the following theorem.

Theorem 4.2. *The optimal solution and the optimal objective value of RA-SP, i.e., Problem (4–2), converge to the optimal solution and the optimal objective value of SP, i.e., Problem (4–1), respectively.*

Proof. According to Proposition 4.4 and Corollary 2, we know as the historical data size M goes to infinity, the true distribution $\mathbb{P}^*$ converges to $\mathbb{P}_0$ in probability. Since as

M goes to infinity, θ goes to zero, so the worst-case probability distribution, denoted as $\mathbb{P}'$, converges to $\mathbb{P}_0$. Therefore, the worst-case distribution $\mathbb{P}'$ converges to the true distribution $\mathbb{P}^*$ in probability. Since Ω is bounded, then for any given $x \in X$, the function $Q(x, \xi)$ is bounded and continuous with ξ. According to the Helly-Bray theorem, we can claim that for any given $x \in X$,

$$\lim_{M\to\infty} \sup_{\mathbb{P}: d(\mathbb{P},\mathbb{P}^0)\leq\theta} E_{\mathbb{P}}[Q(x,\xi)] = \lim_{M\to\infty} E_{\mathbb{P}'}[Q(x,\xi)] = E_{\mathbb{P}^*}[Q(x,\xi)]. \tag{4–23}$$

With equation (4–23), we first explore the convergence property of optimal values. We represent $\hat{v}$ as the optimal value and $\hat{x}$ as the optimal solution of the following problem:

$$\begin{aligned} &\min_x \quad c^T x + \lim_{M\to\infty} \sup_{\mathbb{P}: d(\mathbb{P},\mathbb{P}^0)\leq\theta} E_{\mathbb{P}}[Q(x,\xi)] \\ &s.t. \quad x \in X. \end{aligned} \tag{4–24}$$

Besides, we represent $\bar{v}$ as the optimal value and $\bar{x}$ as the optimal solution of problem SP (i.e., problem as described in (4–1)). Then we need to prove that $\hat{v} = \bar{v}$. Due to the fact that $\hat{v} \geq \bar{v}$, if the equation $\hat{v} = \bar{v}$ does not hold, we have $\hat{v} > \bar{v}$. According to Equation (4–23), we can observe that

$$c^T \bar{x} + \lim_{M\to\infty} \sup_{\mathbb{P}: d(\mathbb{P},\mathbb{P}^0)\leq\theta} E_{\mathbb{P}}[Q(\bar{x},\xi)] = c^T \bar{x} + E_{\mathbb{P}^*}[Q(\bar{x},\xi)]. \tag{4–25}$$

Therefore,

$$\begin{aligned} & c^T \hat{x} + \lim_{M\to\infty} \sup_{\mathbb{P}: d(\mathbb{P},\mathbb{P}^0)\leq\theta} E_{\mathbb{P}}[Q(\hat{x},\xi)] \\ &= \hat{v} \\ &> \bar{v} \\ &= c^T \bar{x} + E_{\mathbb{P}^*}[Q(\bar{x},\xi)] \\ &= c^T \bar{x} + \lim_{M\to\infty} \sup_{\mathbb{P}: d(\mathbb{P},\mathbb{P}^0)\leq\theta} E_{\mathbb{P}}[Q(\bar{x},\xi)], \end{aligned} \tag{4–26}$$

which violates that $\hat{x}$ is the optimal solution to problem (4–24). Consequently, we have $\hat{v} = \bar{v}$, which indicates the convergence property of optimal values. Besides, since the optimal solution of problem (4–2) converges to $\hat{x}$, and $\hat{x}$ is also an optimal solution of problem (4–1) due to $\hat{v} = \bar{v}$, we can claim that the optimal solution of problem (4–2) converges to the optimal solution of problem (4–1). □

Next, we derive a sampling approach to solve RA-SP (i.e., Problem (4–2)) for the case in which the true probability distribution is continuous, with any metric $\mathcal{F}$ in ζ-structure probability metrics class. We denote

$$f(x) = c^T x + \sup_{\mathbb{P}: d_{\mathcal{F}}(\mathbb{P}, \mathbb{P}_0) \leq \theta} E_{\mathbb{P}}[Q(x, \xi)], \tag{4–27}$$

where $\mathbb{P}_0$ is the probability distribution derived from f_n as indicated in (4–15), and θ can be calculated by using Theorem 4.1 and Corollary 3. We consider the following two problems:

$$H^+(x, \epsilon) = c^T x + \sup_{F_N: d_{\mathcal{F}}(F_N, F_N^0) \leq \theta + \epsilon} E_{F_N}[Q(x, \xi)], \tag{4–28}$$

$$H^-(x, \epsilon) = c^T x + \sup_{F_N: d_{\mathcal{F}}(F_N, F_N^0) \leq \theta - \epsilon} E_{F_N}[Q(x, \xi)], \tag{4–29}$$

where F_N^0 is the empirical distribution of N samples $\xi_0^1, \xi_0^2, \cdots, \xi_0^N$ drawn from the reference distribution $\mathbb{P}_0$, and F_N is any discrete distribution of $\xi_0^1, \xi_0^2, \cdots, \xi_0^N$. For notation brevity, let $\beta_{\mathcal{F}}(\epsilon, N)$ represent the convergence rate corresponds to different metric $\mathcal{F}$ obtained by Proposition 4.4 and Corollary 2. We have the following theorem:

Proposition 4.7. *For a given first stage decision $\bar{x}$, $H^-(\bar{x}, \epsilon)$ is a lower bound of $f(\bar{x})$ with probability $1 - \beta_{\mathcal{F}}(\epsilon, N)$ and $H^+(\bar{x}, \epsilon)$ is an upper bound of $f(\bar{x})$ with probability $1 - \beta_{\mathcal{F}}(\epsilon, N)$.*

Proof. For the lower bound part, we first assume that F_N^* is the worst-case distribution of Problem (4–29). Then F_N^* should satisfy the constraint $d_{\mathcal{F}}(F_N^*, F_N^0) \leq \theta - \epsilon$. Also, based on Proposition 4.4 and Corollary 2, we have $d_{\mathcal{F}}(F_N^0, \mathbb{P}_0) \leq \epsilon$ with probability

$1-\beta_{\mathcal{F}}(\epsilon, N)$. Since $d_{\mathcal{F}}$ is a metric, it satisfies the triangle inequality, i.e., $d_{\mathcal{F}}(F_N^*, \mathbb{P}_0) \leq d_{\mathcal{F}}(F_N^*, F_N^0) + d_{\mathcal{F}}(F_N^0, \mathbb{P}_0)$. Thus we have $d_{\mathcal{F}}(F_N^*, \mathbb{P}_0) \leq \theta$ with probability $1-\beta_{\mathcal{F}}(\epsilon, N)$. Therefore,

$$
\begin{aligned}
& Pr\{H^-(\bar{x}, \epsilon) \leq f(\bar{x})\} \\
&= Pr\{c^T\bar{x} + E_{F_N^*}[Q(\bar{x}, \xi)] \leq c^T\bar{x} + \sup_{\mathbb{P}: d_{\mathcal{F}}(\mathbb{P}, \mathbb{P}_0) \leq \theta} E_{\mathbb{P}}[Q(\bar{x}, \xi)]\} \\
&= Pr\{E_{F_N^*}[Q(\bar{x}, \xi)] \leq \sup_{\mathbb{P}: d_{\mathcal{F}}(\mathbb{P}, \mathbb{P}_0) \leq \theta} E_{\mathbb{P}}[Q(\bar{x}, \xi)]\} \\
&\geq Pr\{F_N^* \in \{\mathbb{P} : d_{\mathcal{F}}(\mathbb{P}, \mathbb{P}_0) \leq \theta\}\} \\
&\geq 1-\beta_{\mathcal{F}}(\epsilon, N).
\end{aligned}
$$

By using the same logic, we can prove the upper bound part. The detailed proof is omitted here.

□

From the proposition we can see, on one hand, as ϵ decreases, the lower bound tends to increase, and the upper bound tends to decrease, and the upper bound and lower bound converge as ϵ goes to 0. But notice here when ϵ is very small, we need to find a large N to guarantee a high probability. On the other hand, consider the problem

$$
H(x) = c^T x + \sup_{F_N: d_{\mathcal{F}}(F_N, F_N^0) \leq \theta} E_{F_N}[Q(x, \xi)] \tag{4–30}
$$

and we know $H^+(x, \epsilon) \geq H(x) \geq H^-(x, \epsilon)$. Therefore, $H(x)$ also converges to the objective of RA-SP exponentially fast. Then we can also solve RA-SP by solving the minimization problem of (4–30).

Now we analyze the upper bound and lower bound of problem RA-SP. We let

$$
v_\epsilon^+ = \min_{x \in X} H^+(x, \epsilon), \tag{4–31}
$$

$$
v_\epsilon^- = \min_{x \in X} H^-(x, \epsilon), \tag{4–32}
$$

and let v^* as the optimal value of RA-SP. We have the following theorem:

Theorem 4.3.

$$Pr\{v_\epsilon^- \leq v^*\} \geq 1 - \beta_{\mathcal{F}}(\epsilon, N), \tag{4–33}$$

$$Pr\{v^* \leq v_\epsilon^+\} \geq 1 - \beta_{\mathcal{F}}(\epsilon, N). \tag{4–34}$$

Proof. We just verify the statement (4–33), and then the statement (4–34) holds following the same logic. Denote x^* and x_ϵ^- are the optimal solutions to the problem RA-SP and problem (4–32), respectively.

$$\begin{aligned}
&Pr\{v_\epsilon^- \leq v^*\} \\
&= Pr\{H^-(x_\epsilon^-, \epsilon) \leq f(x^*)\} \\
&\geq Pr\{H^-(x^*, \epsilon) \leq f(x^*)\} && (4–35) \\
&\geq 1 - \beta_{\mathcal{F}}(\epsilon, N), && (4–36)
\end{aligned}$$

where inequality (4–35) holds since x_ϵ^- is the optimal solution to the problem (4–32) so that $H^-(x_\epsilon^-, \epsilon) \leq H^-(x^*, \epsilon)$. The inequality (4–36) is derived by using Proposition 4.7. □

Now we discuss the calculations of θ values for the continuous case. We can follow the similar logic as discussed in Subsection 4.3.1 and represent θ as a function of confidence level η and the number of historical data M, based on Theorem 4.1 and Corollary 3. In addition, if the probabilities for (4–28) to be the upper bound of (4–27) and (4–29) to be the lower bound of (4–27) are given, we can also calculate the number of ϵ, based on Proposition 4.4 and Corollary 2.

Next, we conclude the algorithm to solve the case in which the true distribution is continuous in Algorithm 2.

Note here that for the case that the true distribution is discrete but the supporting space Ω is infinite, we can employ the same algorithm, except using the empirical distribution as the reference distribution. In that case, the convergence rate (i.e., the value of θ) can be obtained by using Proposition 4.4 and Corollary 2.

Algorithm 2 Algorithm for the continuous case

Input: Historical data $\xi^1, \xi^2, \cdots, \xi^M$ drawn i.i.d. from the true distribution and confidence level η and τ.
Output: The estimated objective value of the risk-averse problem.
Step 1: Obtain the kernel density function $f_M(x)$ as shown in (4–15) and θ based on the historical data.
Step 2: Set gap$= 1000$, $N = 0$.
Step 3: While gap $> \sigma$, do
1) $N = N + \Delta N$.
2) Simulate N samples from $f_M(x)$, and obtain ϵ based on Proposition 4.4 and Corollary 2.
3) Obtain the upper bound by solving (4–31).
4) Obtain the lower bound by solving (4–32).
5) Obtain the optimality gap between the lower and upper bounds.
Step 4: Output the solution.

4.4 Numerical Experiments

In this section, by using the ζ-structure probability metrics to construct the confidence set, we explore the performances for both discrete and continuous cases. We show the system performance by two examples: Newsvendor problem and Facility Location problem.

4.4.1 Newsvendor Problem

In this subsection, we use a two-stage Newsvendor problem as an example to conduct our numerical experiments. A news vendor places an order y of newspapers one day ahead at a unit cost of c_1. On the second day, she observes the real demand d and places a supplemental order x at a unit cost of c_2. If $c_1 \geq c_2$, then the news vendor can put all her purchase orders on the second day, based on the known information of demand. For this case, the two-stage newsvendor problem can be reduced to be a single-stage problem. If $c_2 \geq p$, then the news vendor will not put any order on the second day to avoid loss of money, and similarly, the two-stage newsvendor problem can be reduced to be a single-stage problem. Therefore, without loss of generalization, we assume $c_1 < c_2 < p$. Accordingly, the data-driven risk-averse two-stage newsvendor

problem can be formulated as follows:

$$\max_{y\geq 0} -c_1 y + \min_{\mathbb{P}:\, d_\zeta(\mathbb{P},\mathbb{P}_0)\leq\theta} E_{\mathbb{P}} \max_{x\geq 0}[p\min\{x+y,\, d(\xi)\} - c_2 x], \tag{4–37}$$

where ζ can be any metric in the ζ-structure probability metrics class, and θ can be calculated based on Proposition 4.1 and Corollary 2 if the true distribution is discrete and Theorem 4.4 and Corollary 3 if the true distribution is continuous.

In the following, by using the newsvendor problem as an example, we explore the performance of our approach for the discrete case. We set $c_1 = 3, c_2 = 4$, and $p = 5$. We assume the demand follows a discrete distribution with two scenarios: 10 and 20 each with probabilities 0.4 and 0.6 respectively. We first study the effects of the number of observed historical data, by setting the confidence level to be 0.99 and varying the number of historical data from 10 to $10,000$.

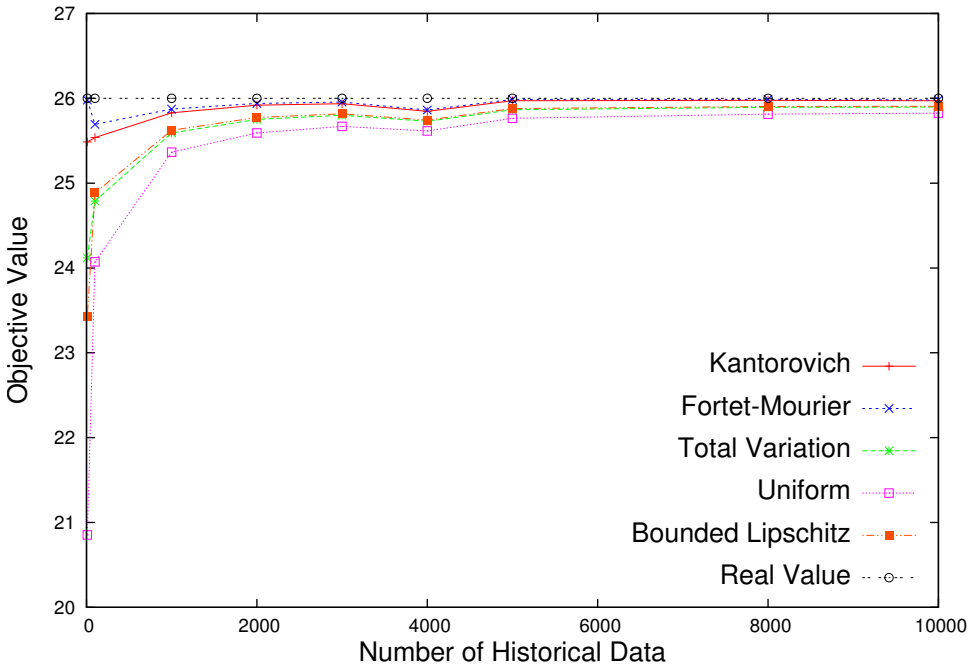

Figure 4-2. Effects of historical data

The results are shown in Figure 4-2. From the figure, we can observe that, no matter what kind of metrics we use, as the number of historical data increases, the objective value of problem RA-SP tends to increase. This result conforms to the

intuition, because the value of θ decreases as the number of historical data increases. Therefore, accordingly, the risk-averse problem RA-SP becomes less conservative. We can also observe that, when the number of historical data exceeds $2,000$, the gaps between the risk-averse problem RA-SP and risk-neutral problem SP are very small (less than 0.2) under the Wasserstein and Fortet-Mourier metrics. Furthermore, when the number of historical data exceeds $5,000$, the gaps between the risk-averse problem RA-SP and the risk-neutral problem SP are small under all metrics.

Next, we explore the effects of confidence level on the objective value of RA-SP. We set the number of historical data to be $5,000$ and test five different confidence levels: 0.7, 0.8, 0.9, 0.95, and 0.99. We report the results in Figure 4-3.

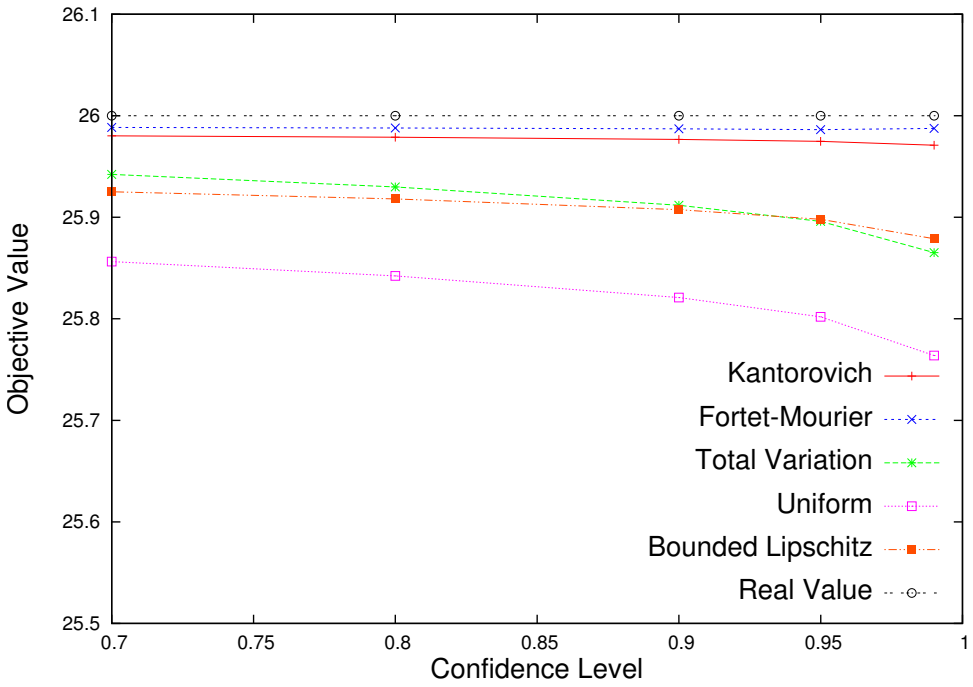

Figure 4-3. Effects of confidence level

From the figure, we can observe that as the confidence level increases, the gaps between the risk-averse problem RA-SP and risk-neutral problem SP also increase. This is due to the fact that, as the confidence level increases, the value of θ increases. Thus, the problem becomes more conservative and the true probability distribution is more likely to be in the confidence set $\mathcal{D}$.

4.4.2 Facility Location Problem

In this subsection, to show the system performance for the continuous distribution case, we use the facility location problem as an example, e.g., a company that supplies independent grocery stores decides what facilities, among locations $i = 1, \cdots, M$, to open. Each facility i associates with a fixed cost F_i and a capacity C_i if it is open. Assume that the company has a stable set of demand sites $j = 1, \cdots, N$, whose demands are independently distributed. There is a shipment cost to bring a unit product from facility i to demand site j. At the first stage, the company needs to decide which facilities to open, and at the second stage, after realizing the demand d_j at site j, the company needs to decide the amount of products to be shipped from facility i to site j. Accordingly, the data-driven risk-averse two-stage stochastic facility location problem can be formulated as follows:

$$
\begin{aligned}
&\min_y \sum_{i=1}^{M} F_i y_i + \max_{\mathbb{P}\in\mathcal{D}} E_{\mathbb{P}}\left[\sum_{i=1}^{M}\sum_{j=1}^{N} T_{ij}x_{ij}(\xi)\right] \\
&s.t. \sum_{j=1}^{N} x_{ij} \le C_i y_i, i = 1, \cdots, M, \\
&\quad \sum_{i=1}^{M} x_{ij} = d_j(\xi), j = 1, \cdots, N, \\
&\quad y_i \in \{0,1\}, x_{ij} \ge 0, i = 1, \cdots, M, j = 1, \cdots, N.
\end{aligned}
$$

We assume there are 10 facility locations and 10 demand sites. For each location i, the capacity is $15 + i$ and the fixed cost is $100 + i$. Meanwhile, the unit shipping cost from location i to demand site j is $5 + 0.008i$. In this experiment, we test the cases in which the random variable follows a Uniform distribution, Normal distribution, Gamma distribution and Weibull distribution, respectively.

Similar to the discrete case, we study the effect of historical data first. We set the number of samples we take from the reference distribution as 100, and compute the

estimated value of the risk averse problem (4–2) described as (4–30). We report the objective values corresponding to various numbers of historical data in Figure 4-5.

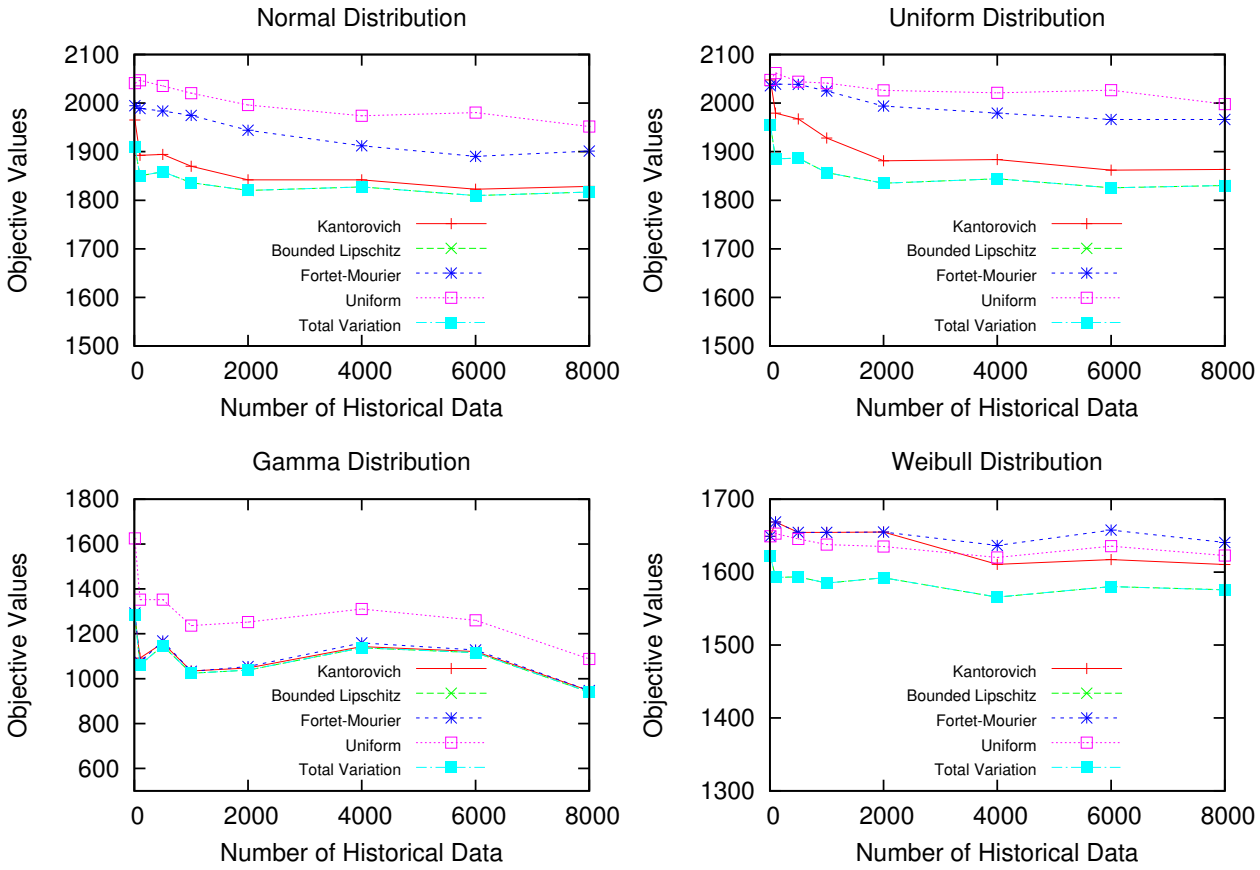

Figure 4-4. Effects of historical data

From Figure 4-4 we can observe that, for a fixed number of samples, as the number of historical data increases, the objective values tend to decrease. That is because as the number of historical data increases, the value of θ decreases and the problem becomes less conservative. In addition, we report the facilities that are not open in the first stage in Table 4.4.2 for the Weibull metric for comparison. From the table we can observe that, for different metrics and different number of historical data, we have different first-stage decisions.

Next, we set the number of historical data to be 5000, and test the effects of the number of samples for the proposed sampling approach. We obtain the upper and

Metrics	10	100	500	1000	2000	4000	6000	8000
K	4,10	1,10	3,10	3,10	3,10	6,10	3,10	6,10
BL	4,10	1,10	3,10	3,10	3,10	7,9	3,10	6,10
FM	4,10	1,10	3,10	3,10	3,10	6,10	3,10	6,10
U	4,10	1,10	3,10	4,9	4,9	7,9	3,10	7,9
TV	5,9	1,10	4,9	4,9	3,10	6,10	3,10	7,9

Table 4-1. Facilities that are not open

lower bounds, and the estimated values corresponding to different sample numbers respectively and show the results in Figure 4-5.

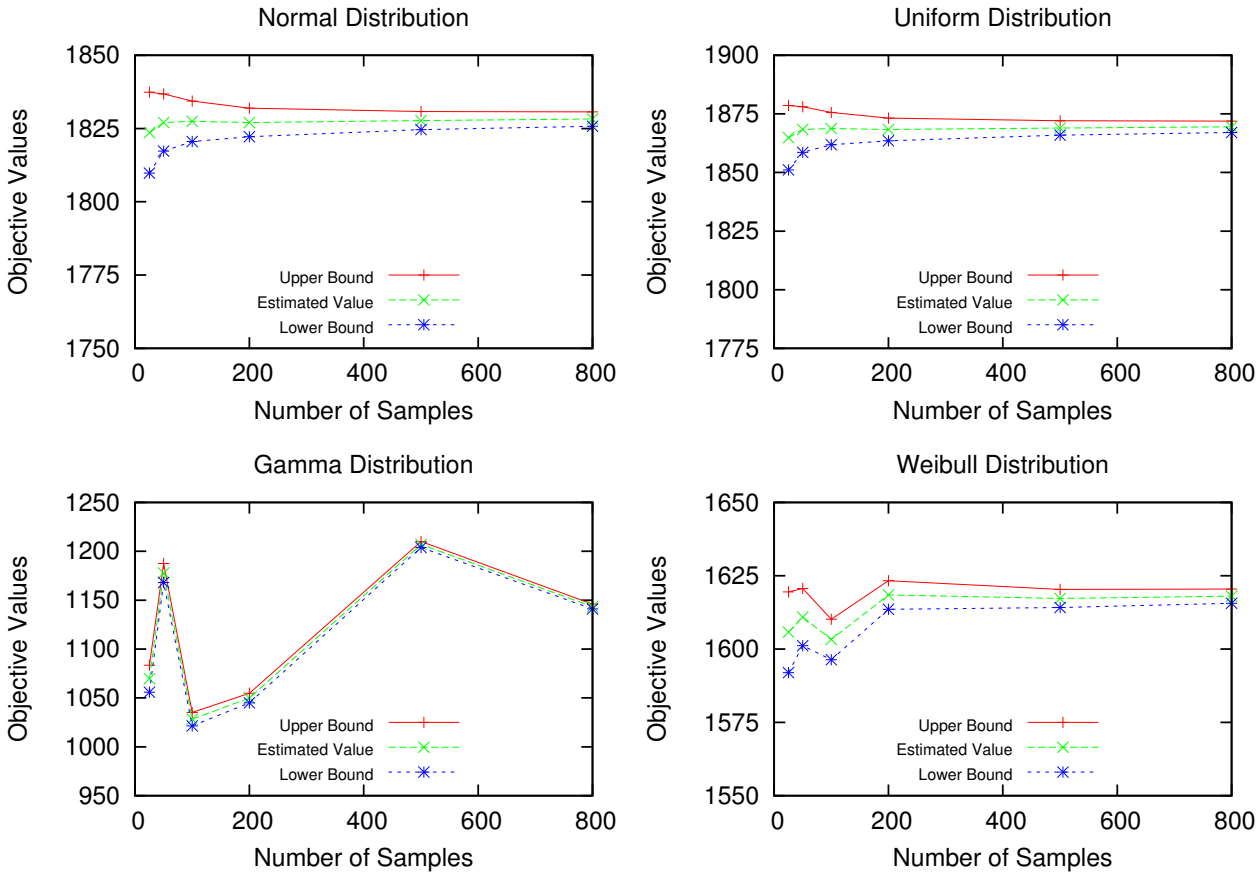

Figure 4-5. Effects of samples

From Figure 4-5 we can observe, no matter what the true distribution is, as the number of samples increases, the gap between the upper bound and the lower bound tends to decrease.

4.5 Summary

In this chapter, we apply a new class of probability metrics, ζ-structure probability metrics, to construct the confidence set of ambiguous distributions, by learning from the historical data. Based on this, we further develop a framework to solve the risk-averse two-stage stochastic program for both cases in which the true distribution is discrete and continuous, respectively. We reformulate the risk-averse problem as a traditional robust optimization problem for the discrete case. For the continuous case, we propose a re-sampling approach to provide the statistical upper and lower bounds for the optimal objective value of the risk-averse problem. In addition, these bounds are proved to converge to the optimal objective value as the sample size increases to infinity. We also prove that under ζ-structure probability metrics, the risk-averse problem converges to the risk-neutral one exponentially fast as the number of historical data increases to infinity. The experimental results of newsvendor and facility location problems show the effectiveness of the proposed approach and numerically show the value of data.

CHAPTER 5
DATA-DRIVEN UNIT COMMITMENT PROBLEM

5.1 Problem Description and Literature Review

In this chapter, we present a data-driven risk-averse stochastic optimization approach to solve the unit commitment problem. That is, instead of knowing the exact distribution of the uncertain parameter, we can observe a series of historical data which are drawn from the true distribution. By learning from the historical data, we construct a reference distribution of the uncertain parameter as well as a confidence set of the true distribution, and make decisions in consideration of the case that the true distribution can vary within the confidence set. Then, we develop a data-driven risk-averse two-stage stochastic unit commitment framework and propose the solution approaches to solve the developed framework. Our contributions can be described as follows:

1. Unlike the traditional two-stage stochastic unit commitment problem, we do not assume that the probability distributions of the uncertain parameters, e.g., electricity load and wind power output, are known. Instead, we construct confidence sets of the probability distributions and propose the solution methodology to solve a risk-averse stochastic unit commitment problem.

2. We examine the convergence properties of the data-driven stochastic unit commitment framework, and we demonstrate that as the number of historical data increases, the risk-averse problem converges to the risk-neutral problem, i.e., the traditional two-stage stochastic unit commitment problem. In that case, our problem becomes less conservative as more historical data are observed.

The remainder of this chapter is organized as follows. In Section 5.2, we describe the mathematical formulations of the traditional stochastic unit commitment problem and the data-driven stochastic unit commitment problem respectively. In Section 5.3, we discuss the solution methodologies to solve the data-driven stochastic unit commitment problem. In Section 5.4, we prove that as the number of historical data increases, the data-driven stochastic unit commitment problem converges to the traditional stochastic unit commitment problem. In Section 5, we conclude our work.

5.2 Mathematical Formulations

In this section, we define the notations and introduce the formulation of the traditional two-stage stochastic unit commitment problem first. Then based on the observed historical data, we construct the reference distribution and the confidence set of the true probability distribution, by bringing the concept of "Wasserstein metric". Based on the reference distribution and the confidence set, we then build the formulation of the data-driven risk-averse two-stage stochastic unit commitment problem.

5.2.1 Stochastic Unit Commitment Problem

In this chapter, we have two dimensions of uncertainties, i.e., wind power output and electricity load. However, since the wind power output can be regarded as a negative load, without loss of generality, we only consider the electricity load as the uncertain parameter. We summarize the notations by sets, parameters, first-stage variables and second-stage variables listed as follows.

A. Sets and Parameters

B Index set of all buses.

$\mathcal{E}$ Index set of transmission lines linking two buses.

G_b Set of thermal generators at bus b.

T Time horizon (e.g., 24 hours).

SU_i^b Start-up cost of thermal generator i at bus b.

SD_i^b Shut-down cost of thermal generator i at bus b.

$F_i(.)$ Fuel cost of thermal generator i.

MU_i^b Minimum up-time for thermal generator i at bus b.

MD_i^b Minimum down-time for thermal generator i at bus b.

RU_i^b Ramp-up rate limit for thermal generator i at bus b.

RD_i^b Ramp-down rate limit for thermal generator i at bus b.

L_i^b Lower bound of electricity generated by thermal generator i at bus b.

U_i^b Upper bound of electricity generated by thermal generator i at bus b.

C_{ij} Capacity for the transmission line linking bus i and bus j.

K_{ij}^b Line flow distribution factor for the transmission line linking bus i and bus j, due to the net injection at bus b.

D_t^{b+} The upper bound of the load at bus b in time t.

D_t^{b-} The lower bound of the load at bus b in time t.

$d_{tb}(\xi)$ The load at bus b in time t corresponding to scenario ξ.

γ_{it}^{jb} The intercept of the jth segment line for the generation cost for generator i at bus b in time t.

β_{it}^{jb} The slope of the jth segment line for the generation cost for generator i at bus b in time t.

B. First-stage Variables

y_{it}^b Binary decision variable: "1" if thermal generator i at bus b is on in time t; "0" otherwise.

u_{it}^b Binary decision variable: "1" if thermal generator i at bus b is started up in time t; "0" otherwise.

v_{it}^b Binary decision variable: "1" if thermal generator i at bus b is shut down in time t; "0" otherwise.

C. Second-stage Variables

$q_{it}^b(\xi)$ Electricity generation amount by thermal generator i at bus b in time t corresponding to scenario ξ.

$\eta_{it}^b(\xi)$ Auxiliary variable representing the fuel cost of thermal generator i at bus b in time t corresponding to scenario ξ.

With the notations, we develop a two-stage stochastic unit commitment formulation by minimizing the expected total generation cost. The first stage is to obtain the day-ahead on/off schedule (i.e., unit commitment decision) for the thermal generators adhering to the operational constraints. After the uncertain load is realized, the second stage is to obtain the real time economic dispatch decision for the thermal generators while satisfying the physical and transmission constraints. The detailed formulation (denoted

as SUC) can be described as follows:

$$\min \sum_{t=1}^{T} \sum_{b \in B} \sum_{i \in G_b} (SU_i^b u_{it}^b + SD_i^b v_{it}^b) + E[\mathcal{Q}(y, u, v, \xi)] \tag{5–1}$$

$s.t.$

$$-y_{i(t-1)}^b + y_{it}^b - y_{ik}^b \leq 0,$$

$$\forall k : 1 \leq k - (t-1) \leq MU_i^b, \quad \forall i \in G_b, \forall b \in B, \forall t, \tag{5–2}$$

$$y_{i(t-1)}^b - y_{it}^b + y_{ik}^b \leq 1,$$

$$\forall k : 1 \leq k - (t-1) \leq MD_i^b, \quad \forall i \in G_b, \forall b \in B, \forall t, \tag{5–3}$$

$$-y_{i(t-1)}^b + y_{it}^b - u_{it}^b \leq 0, \quad \forall i \in G_b, \forall b \in B, \forall t, \tag{5–4}$$

$$y_{i(t-1)}^b - y_{it}^b - v_{it}^b \leq 0, \quad \forall i \in G_b, \forall b \in B, \forall t, \tag{5–5}$$

$$y_{it}^b, u_{it}^b, v_{it}^b \in \{0, 1\}, \quad \forall i \in G_b, \forall b \in B, \forall t, \tag{5–6}$$

where $\mathcal{Q}(y, u, v, \xi)$ is equal to

$$\min \sum_{t=1}^{T} \sum_{b \in B} \sum_{i \in G_b} F_i(q_{it}^b(\xi)) \tag{5–7}$$

$s.t.$

$$L_i^b y_{it}^b \leq q_{it}^b(\xi) \leq U_i^b y_{it}^b, \quad \forall i \in G_b, \forall b \in B, \forall t, \tag{5–8}$$

$$q_{it}^b(\xi) - q_{i(t-1)}^b(\xi) \leq (2 - y_{i(t-1)}^b - y_{it}^b) L_i^b +$$

$$(1 + y_{i(t-1)}^b - y_{it}^b) RU_i^b, \quad \forall i \in G_b, \forall b \in B, \forall t, \tag{5–9}$$

$$q_{i(t-1)}^b(\xi) - q_{it}^b(\xi) \leq (2 - y_{i(t-1)}^b - y_{it}^b) L_i^b +$$

$$(1 - y_{i(t-1)}^b + y_{it}^b) RD_i^b, \quad \forall i \in G_b, \forall b \in B, \forall t, \tag{5–10}$$

$$-C_{ij} \leq \sum_{b \in B} K_{ij}^b (\sum_{r \in G_b} q_{rt}^b(\xi) - d_{tb}(\xi)) \leq C_{ij}, \quad \forall (i, j) \in \mathcal{E}, \forall t, \tag{5–11}$$

$$\sum_{b \in B} \sum_{i \in G_b} q_{it}^b(\xi) = \sum_{b \in B} d_{tb}(\xi), \quad \forall t. \tag{5–12}$$

In the above formulation, constraints (5–2) and (5–3) indicate that a minimum up-time and a minimum down-time are needed in order to start up and shut down the thermal unit. Constraints (5–4) and (5–5) are the start-up and shut-down operational constraints

for each thermal unit. Constraint (5–8) enforces the upper and lower bounds of the electricity generation amount of each thermal unit. The ramping up constraint (5–9) and ramping down constraint (5–10) indicate the maximum increment and decrement of the power generation amount of each unit between two adjacent periods when the thermal unit is on. Constraint (5–11) represents the transmission capacity constraints and constraint (5–12) ensures load balance.

Next, we introduce the traditional solution approach to address the stochastic unit commitment problem. First, note that the fuel cost function $F_i(.)$ is normally a quadratic function. The traditional way to linearize the quadratic function is to use multiple pieces of piecewise linear function to approximate it, and we also utilize this technique in this chapter. For instance, we use a K-piece piecewise linear function as follows to approximate $F_i(.)$:

$$\eta_{it}^b \geq \gamma_{it}^{jb} y_{it}^b + \beta_{it}^{jb} x_{it}^b, \tag{5–13}$$
$$\forall t = 1, \cdots, T, \forall b \in B, \forall i \in G_b, \forall k = 1, \cdots, K.$$

In order to keep notations brevity, we use the abstract formulation to represent SUC. For instance, we use vectors to represent variables and matrix to represent constraints, then SUC can be abstracted as follows:

$$\begin{aligned} &\min_{y,u,v} (\mathbf{a}^T u + \mathbf{b}^T v) + E[\mathcal{Q}(y, u, v, \xi)] \\ &s.t. \quad \mathbf{A}y + \mathbf{B}u + \mathbf{C}v \geq \mathbf{r}, \end{aligned} \tag{5–14}$$

where $\mathcal{Q}(y, u, v, \xi)$ is equal to

$$\min \ \mathbf{e}^T \eta(\xi) \tag{5–15}$$
$$s.t. \ \mathbf{D}x(\xi) \geq \mathbf{f} + \mathbf{F}y, \tag{5–16}$$
$$\mathbf{G}x(\xi) \geq \mathbf{g} + \mathbf{H}d(\xi), \tag{5–17}$$
$$\mathbf{J}x(\xi) + \mathbf{K}\eta(\xi) \geq \mathbf{h} + \mathbf{L}y. \tag{5–18}$$

Note here that constraint (5–14) represents constraints (5–2) - (5–5); constraint (5–16) represents constraints (5–8) - (5–11); constraint (5–17) represents constraint (5–12) and constraint (5–18) represents constraint (5–13).

To solve SUC, scenario-based approaches are commonly used. The key idea of scenario-based approaches is to assume that the random parameter follows a given distribution, for example, multivariate normal distribution $N(D, \Sigma)$ with its predicted value D and volatility matrix Σ. In addition, simulation techniques such as Monto Carlo simulation are commonly used to generate scenarios for the random parameter. Therefore, the expected cost in the objective function can be estimated as the average value of costs corresponding to different generated scenarios. For example, if we use Monto Carlo simulation to generate N scenarios for the electricity load, i.e., $d_{tb}(\xi^1), d_{tb}(\xi^2), \cdots, d_{tb}(\xi^N)$, we can approximate the objective function of SUC as follows:

$$\min_{y,u,v}(\mathbf{a}^T u + \mathbf{b}^T v) + \frac{1}{N}\sum_{n=1}^{N} \mathcal{Q}(y, u, v, \xi^n). \tag{5–19}$$

Accordingly, the stochastic unit commitment problem is transformed into a deterministic MIP to a larger extent. Based on the lagrange relaxation technique, the problem can be decomposed into scenario-based subproblems [53], and then can be addressed by solving each subproblem.

However, due to the incomplete information of the electricity load, it is very difficult to accurately estimate its probability distribution. Consequently, the obtained unit commitment decision can be biased. For example, if the online units are not capable of providing enough electricity to balance the load, electricity shortage or blackout may occur. On the other side, if the generated electricity exceeds the electricity consumption, the wind power output needs to be curtailed to keep the system balanced. To tackle these challenging problems, the concept of data-driven risk-averse stochastic optimization has been proposed. In the next section, we introduce the data-driven

risk-averse two-stage stochastic unit commitment problem and discuss the solution approaches to address the problem.

5.2.2 Data-Driven Unit Commitment Formulation

In this section, we develop a data-driven risk-averse two-stage stochastic unit commitment model, in which instead of knowing the exact distribution of the electricity load, a series of historical data of load are observed. Based on the observed historical data, we build the confidence set for the probability distribution of the electricity load, by using the Wasserstein metric. Then we propose the solution methodology to solve the data-driven risk-averse two-stage stochastic unit commitment problem. Finally, we analyze the convergence properties of the model and show that the risk-averseness of the model decreases as more historical data are observed and eventually does not exist as the number of historical data goes to infinity.

5.2.2.1 Wasserstein metric

The Wasserstein metric is defined as a distance function between two probability distributions on a given supporting space Ω. More specifically, given two probability distributions $\mathbb{P}$ and $\mathbb{Q}$ on the supporting space Ω, Wasserstein metric is defined as $d_W(\mathbb{P}, \mathbb{Q}) := \inf_\pi\{E_\pi[\rho(X, Y)] : \mathbb{P} = \mathcal{L}(X), \mathbb{Q} = \mathcal{L}(Y)\}$, where $\rho(X, Y)$ is defined as the distance between random variables X and Y, and the infimum is taken over all joint distributions π with marginals $\mathbb{P}$ and $\mathbb{Q}$. First, as indicated in [68], the Wasserstein metric is indeed a metric since it satisfies the properties of metrics. That is, $d_W(\mathbb{P}, \mathbb{Q}) = 0$ if and only if $\mathbb{P} = \mathbb{Q}$, $d_W(\mathbb{P}, \mathbb{Q}) = d_W(\mathbb{Q}, \mathbb{P})$ (symmetric property) and $d_W(\mathbb{P}, \mathbb{Q}) \leq d_W(\mathbb{P}, \mathcal{O}) + d_W(\mathcal{O}, \mathbb{Q})$ for any probability distribution $\mathcal{O}$ (triangle equality). In addition, by the Kantorovich-Rubinstein theorem [33], the Wasserstein metric is equivalent to the Kantorovich metric, which is defined as

$$d_K(\mathbb{P}, \mathbb{Q}) = \sup_{h \in \mathcal{H}} \left| \int_\Omega h d\mathbb{P} - \int_\Omega h d\mathbb{Q} \right|, \tag{5–20}$$

where $\mathcal{H} = \{h : \|h\|_L \leq 1\}$, and $\|h\|_L := \sup\{\frac{h(x)-h(y)}{\rho(x,y)} : x \neq y \text{ in } \Omega\}$. In particular, when $\Omega = R$,

$$d_W(\mathbb{P}, \mathbb{Q}) = \int_{-\infty}^{+\infty} |F(x) - G(x)| dx, \tag{5–21}$$

where $F(x)$ and $G(x)$ are the distribution functions derived by $\mathbb{P}$ and $\mathbb{Q}$ respectively. This conclusion holds following the argument that $\inf_\pi\{E_\pi[\rho(X, Y)] : \mathbb{P} = \mathcal{L}(X), \mathbb{Q} = \mathcal{L}(Y)\} = \int_0^1 |F^{-1}(t) - G^{-1}(t)| dt$, as stated in Theorem 6.0.2 in [3] and $\int_0^1 |F^{-1}(t) - G^{-1}(t)| dt = \int_{-\infty}^{+\infty} |F(x) - G(x)| dx$. The concept of Wasserstein metric is first introduced by Leonid Vasershtein, and commonly applied in many areas. For example, many metrics known in statistics, measure theory, ergodic theory, functional analysis, etc., are special cases of the Wasserstein / Kantorovich metric [67]. Wasserstein / Kantorovich also has many applications in transportation theory [52], and some applications in computer science like probabilistic concurrency, image retrieval, data mining, and bioinformatics, etc [20]. In this chapter, we use Wasserstein metric to construct the confidence set for the probability distribution.

5.2.2.2 Reference distribution

After observing a series of historical data, we can have an estimation of the true probability distribution, which is called reference probability distribution. Intuitively, the more historical data we have, the more accurate estimation of the true probability distribution we can obtain. Many significant works have been made to get the reference distribution, with both parametric approaches and nonparametric approaches. For instance, for the parametric approaches, the true probability distribution is usually estimated as a particular distribution function, e.g., normal distribution, and the parameters (e.g., mean and variance) are estimated by learning from the historical data. On the other hand, the nonparametric estimations, such as kernel density estimation, are also proved to be effective approaches to obtain the reference probability distribution (e.g., [54] and [51]). In this chapter, we use a nonparametric approach, more specifically,

the empirical distribution function to estimate the true probability distribution. The empirical distribution function is a step function that jumps up by $1/N$ at each of the N independent and identically-distributed data points. That is, given N i.i.d. historical data points $\xi_0^1, \xi_0^2, \cdots, \xi_0^N$, the empirical distribution is defined as

$$\mathbb{P}_0(x) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_0^i}(x), \tag{5–22}$$

where $\delta_{\xi_0^i}(x)$ is one if $x \geq \xi_0^i$ and zero elsewhere. Based on the strong law of large numbers, it can be proved that the reference distribution $\mathbb{P}_0$ pointwise converges to the true probability distribution $\mathbb{P}$ almost surely [66]. By Glivenko-Cantelli theorem, this result can be strengthened by proving the uniform convergence of $\mathbb{P}$ to $\mathbb{P}_0$ [74]. In Chapter 4, we have proved that under the Wasserstein metric, the empirical distribution $\mathbb{P}_0$ exponentially converges to the true distribution $\mathbb{P}$.

5.2.2.3 Confidence set construction

With the previously defined probability metric and reference probability distribution, we now construct the confidence set of the true probability distribution. Intuitively, the more historical data that can be observed, the "closer" the reference distribution is to the true distribution. If we use θ to represent the distance between the reference distribution and the true distribution, then the more historical data we have, the smaller the value of θ is, and the tighter the confidence set becomes. Therefore, the confidence set $\mathcal{D}$ can be represented as follows:

$$\mathcal{D} = \{\mathbb{P} : d(\mathbb{P}, \mathbb{P}_0) \leq \theta\}, \tag{5–23}$$

where the value of θ depends on the number of historical data. More specifically, from Chapter 4, we have shown the exact relationship between the number of historical data and the value of θ, in the following proposition:

Proposition 5.1. *For a general dimension d of the supporting space Ω, we have*

$$P(d_W(\mathbb{P}_0, \mathbb{P}) \leq \theta) \geq 1 - \exp(-\frac{\theta^2}{2B^2}N),$$

where N is the number of historical data, and B is the diameter of Ω. Specifically, if $d = 1$, we have

$$Pr(d_W(\mathbb{P}_0, \mathbb{P}) \leq \theta) \geq 1 - 2\exp(-2N\theta^2).$$

Moreover, we consider a special case that $\mathbb{P}_0$ and $\mathbb{P}$ are product measures. That is, $\mathbb{P}_0 = \otimes_{i=1}^{d}\mathbb{P}_0^{(i)}$ and $\mathbb{P} = \otimes_{i=1}^{d}\mathbb{P}^{(i)}$, where $\mathbb{P}_0^{(i)}$ and $\mathbb{P}^{(i)}$ are the probability distributions defined on the Borel σ-algebra of $\mathbb{R}$. Besides, we consider the case that $\rho(x, y)$ is with $\mathcal{L}_1$ norm (i.e., $\rho(x, y) = \sum_{i=1}^{d}|x_i - y_i|$). Then, based on the definition of Wasserstein metric, we have

$$d_W(\mathbb{P}_0, \mathbb{P}) = \sum_{i=1}^{d} d_W(\mathbb{P}_0^{(i)}, \mathbb{P}^{(i)}),$$

where $d_W(\mathbb{P}_0^{(i)}, \mathbb{P}^{(i)}) = \int_R |\mathbb{P}_0^{(i)}(x) - \mathbb{P}^{(i)}(x)|dx$. Furthermore, according to Hölder's inequality, we have $\int_R |\mathbb{P}_0^{(i)}(x) - \mathbb{P}^{(i)}(x)|dx \leq \|\mathbb{P}_0^{(i)}(x) - \mathbb{P}^{(i)}(x)\|_\infty$. Finally, the Dvoretzky-Kiefer-Wolfowitz theorem leads to the following inequalities:

$$Pr(d_W(\mathbb{P}_0, \mathbb{P}) \geq \theta) \leq \sum_{i=1}^{d} Pr(d_W(\mathbb{P}_0^i, \mathbb{P}^i) \geq \frac{\theta}{d}) \leq 2d\exp(-\frac{2n\theta^2}{d^2}).$$

In conclusion, for the special case that $\mathbb{P}_0$ and $\mathbb{P}$ are product measures and $\rho(x, y) = \sum_{i=1}^{d}|x_i - y_i|$, we have

$$Pr(d_W(\mathbb{P}_0, \mathbb{P}) \leq \theta) \geq 1 - 2d\exp(-\frac{2n\theta^2}{d^2}),$$

for a general $d \geq 1$.

Instead of knowing the true probability distribution $\mathbb{P}$, we assume $\mathbb{P}$ can vary within the confidence set $\mathcal{D}$ and we consider the worst-case expected value $E[\mathcal{Q}(y, u, v, \xi)]$. In that case, the proposed approach is more conservative than the traditional stochastic

optimization approach. That is, the proposed approach is a risk-averse approach. Therefore, the data-driven risk-averse two-stage stochastic unit commitment problem can be formulated as follows:

$$\begin{aligned} &\min_{y,u,v} \ (\mathbf{a}^T u + \mathbf{b}^T v) + \max_{\mathbb{P}\in\mathcal{D}} E_{\mathbb{P}}[\mathcal{Q}(y, u, v, \xi)] \\ (DD-SUC) \ \ &s.t. \ \ \mathbf{A}y + \mathbf{B}u + \mathbf{C}v \geq \mathbf{r}. \end{aligned} \tag{5–24}$$

5.3 Solution Methodology

In this section, we propose the solution approach to address the problem (DD-SUC). We assume the electricity load for each time period t at each bus b is between a lower bound D_t^{b-} and an upper bound D_t^{b+}, and without loss of generality, we let $d(\xi) = \xi$. Accordingly, the supporting space (or uncertainty set) Ω for the random electricity load can be described as follows:

$$\Omega := \left\{ \xi \in \mathcal{R}^{|B|\times|T|} : D_t^{b-} \leq \xi_{tb} \leq D_t^{b+}, \forall t, \forall b \right\}. \tag{5–25}$$

As indicated in Subsection 5.2.2.2, if we have N samples of historical data $\xi^1, \xi^2, \cdots, \xi^N$, the reference distribution $\mathbb{P}_0$ can be defined as the empirical distribution, i.e., $\mathbb{P}_0 = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi^i}(x)$. On the other hand, according to the definition of the Wasserstein metric, the confidence set $\mathcal{D}$ can be reformulated as follows:

$$\mathcal{D} = \left\{ \mathbb{P} : \inf_{\pi} \{ E_\pi[\rho(Z, W)] : \mathbb{P} = \mathcal{L}(Z), \mathbb{P}_0 = \mathcal{L}(W) \} \leq \theta \right\}. \tag{5–26}$$

Therefore, based on the definition of $E_\pi[\rho(Z, W)]$ and properties of conditional density, we can obtain the following reformulation of $E_\pi[\rho(Z, W)]$:

$$E_\pi[\rho(Z, W)] = \frac{1}{N}\sum_{i=1}^{N} \int_{w\in\Omega} f_{w|\xi^i}(w|z = \xi^i)\rho(w, \xi^i)\, dw, \tag{5–27}$$

where $f_{w|\xi^i}(w|z = \xi^i)$ is the conditional density function when $z = \xi^i$. For notation brevity, we let $f^i(w) = f_{w|\xi^i}(w|z = \xi^i)$ and $\rho^i(w) = \rho(w, \xi^i)$. Then the second stage problem of

(DD-SUC) can be reformulated as:

$$\begin{aligned}
&\max_{f_Y(w)\geq 0} \int_{w\in\Omega} \mathcal{Q}(y,u,v,w) f_Y(w)\, dw \\
&s.t. \qquad f_Y(w) = \frac{1}{N}\sum_{i=1}^{N} f^i(w), && \text{(5–28)}\\
&\int_{w\in\Omega} f^i(w)\, dw = 1, \quad \forall i, && \text{(5–29)}\\
&\frac{1}{N}\sum_{i=1}^{N} \int_{w\in\Omega} f^i(w)\rho^i(w)\, dw \leq \theta, && \text{(5–30)}
\end{aligned}$$

where $f_Y(w)$ is the density function of $\mathbb{P}$. The constraints (5–28) and (5–29) are based on the properties of conditional density function and the constraint (5–30) is the representation of the confidence set $\mathcal{D}$. By substituting constraint (5–28) into the objective function, we can obtain its equivalent formulation as follows:

$$\begin{aligned}
&\max_{f^i(w)\geq 0} \frac{1}{N}\sum_{i=1}^{N} \int_{w\in\Omega} \mathcal{Q}(y,u,v,w) f^i(w)\, dw \\
&s.t. \qquad \int_{w\in\Omega} f^i(w)\, dw = 1, \quad \forall i, && \text{(5–31)}\\
&\frac{1}{N}\sum_{i=1}^{N} \int_{w\in\Omega} f^i(w)\rho^i(w)\, dw \leq \theta. && \text{(5–32)}
\end{aligned}$$

Since the above problem is a semi-infinite problem, there is no duality gap. Then we can consider the Lagrangian dual problem that can be written as:

$$L(\lambda_i,\beta) = \max_{f^i\geq 0} \frac{1}{N}\sum_{i=1}^{N} \int (\mathcal{Q}(y,u,v,w) - N\lambda_i - \beta\rho^i(w)) f^i(w)\, dw + \sum_{i=1}^{N} \lambda_i + \theta\beta, \quad \text{(5–33)}$$

where λ_i and β are dual variables of constraints (5–31) and (5–32) respectively. The dual problem is then to minimize $L(\lambda_i,\beta)$ with respect to λ_i, β, and β is subject to be nonnegative. Next, we argue that $\forall w \in \Omega$, $\mathcal{Q}(y,u,v,w) - N\lambda_i - \beta\rho^i(w) \leq 0$. If this argument does not hold, then there exists w_0 such that $\mathcal{Q}(y,u,v,w_0) - N\lambda_i - \beta\rho^i(w_0) > 0$. It means there exists a small number σ, such that $\mathcal{Q}(y,u,v,w_0) - N\lambda_i - \beta\rho^i(w_0) > \sigma$. Since the function $\mathcal{Q}(y,u,v,w) - N\lambda_i - \beta\rho^i(w)$ is continuous with w, if $\mathcal{Q}(y,u,v,w_0) -$

$N\lambda_i - \beta\rho^i(w_0) > \sigma$, there exists a small ball $B(w_0, \epsilon)$, such that $\mathcal{Q}(y, u, v, w) - N\lambda_i - \beta\rho^i(w) > \sigma$ for $\forall w \in B(w_0, \epsilon)$. Therefore, we can let $f^i(w)$ to be arbitrary large when $w \in B(w_0, \epsilon)$, then $L(\lambda_i, \beta)$ is arbitrary large as well, which leads to a contradiction. Hence, the argument $\mathcal{Q}(y, u, v, w_0) - N\lambda_i - \beta\rho^i(w_0) \leq 0$ for all $w \in \Omega$ holds. In that case,

$$\max_{f^i \geq 0} \frac{1}{N} \sum_{i=1}^{N} \int (\mathcal{Q}(y, u, v, w_0) - N\lambda_i - \beta\rho^i(w_0)) f^i(w) dw + \sum_{i=1}^{N} \lambda_i + \theta\beta = \sum_{i=1}^{N} \lambda_i + \theta\beta,$$

with the optimal solution $f^i = 0, i = 1, \cdots, N$. Then, the dual formulation is reformulated as:

$$\begin{aligned} &\min_{\lambda_i, \beta \geq 0} \sum_{i=1}^{N} \lambda_i + \theta\beta \\ &s.t. \quad \mathcal{Q}(y, u, v, w) - N\lambda_i - \beta\rho^i(w) \leq 0, \ \forall w \in \Omega, \forall i = 1, \cdots, N. \end{aligned}$$

From the above formulation, it is easy to observe that the optimal λ_i is equal to $\frac{1}{N} \max_{w \in \Omega} \{\mathcal{Q}(y, u, v, w) - \beta\rho^i(w)\}$ and the second-stage risk-averse optimization problem is equivalent to

$$\min_{\beta \geq 0} \left\{ \frac{1}{N} \sum_{i=1}^{N} \max_{w \in \Omega} \{\mathcal{Q}(y, u, v, w) - \beta\rho^i(w)\} + \theta\beta \right\}. \tag{5–34}$$

Therefore, we have the following theorem:

Theorem 1. *The problem (DD-SUC) is equivalent to the following two-stage robust optimization problem:*

$$\min_{y, u, v, \beta \geq 0} (\boldsymbol{a}^T u + \boldsymbol{b}^T v) + \theta\beta + \frac{1}{N} \sum_{i=1}^{N} \max_{w \in \Omega} \{\mathcal{Q}(y, u, v, w) - \beta\rho^i(w)\} \tag{5–35}$$

$$(R - SUC) \ s.t. \ \ \boldsymbol{A}y + \boldsymbol{B}u + \boldsymbol{C}v \geq \boldsymbol{r}, \tag{5–36}$$

where $\mathcal{Q}(y, u, v, w)$ is equal to

$$\min \ \ \boldsymbol{e}^T \eta(w) \tag{5–37}$$

$$s.t. \ \ \boldsymbol{D}x(w) \geq \boldsymbol{f} + \boldsymbol{F}y, \tag{5–38}$$

$$\boldsymbol{G}x(w) \geq \boldsymbol{g} + \boldsymbol{H}w, \tag{5–39}$$

$$\boldsymbol{J}x(w) + \boldsymbol{K}\eta(w) \geq \boldsymbol{h} + \boldsymbol{L}y. \tag{5–40}$$

Next, we discuss the solution approach to address the (R-SUC) problem. In this chapter, we utilize the Benders' decomposition algorithm to solve the problem. First, we dualize the constraints (5–38) to (5–40) and obtain the following dual formulation and combine it with the second stage to get the subproblem (denoted as SUB-SUC):

$$\begin{aligned}
\phi^i(y) = \max_{w,\lambda,\mu,\nu} \; & (\mathbf{f} + \mathbf{F}y)^T\lambda + (\mathbf{g} + \mathbf{H}w)^T\mu + (\mathbf{h} + \mathbf{L}y)^T\nu - \beta\rho^i(w) \\
s.t. \quad & \mathbf{D}^T\lambda + \mathbf{G}^T\mu + \mathbf{J}^T\mu \leq 0, && (5–41) \\
& \mathbf{K}^T\mu = \mathbf{e}, && (5–42) \\
& \lambda, \mu, \nu \geq 0, && (5–43)
\end{aligned}$$

where λ, μ, ν are dual variables for constraints (5–38), (5–39) and (5–40) respectively. Let π^i denote the optimal value of the subproblem, we can obtain the following master problem:

$$\begin{aligned}
\min_{y,u,v} \; & (\mathbf{a}^T u + \mathbf{b}^T v) + \frac{1}{N}\sum_{i=1}^{N} \pi^i \\
s.t. \quad & \mathbf{A}y + \mathbf{B}u + \mathbf{C}v \geq \mathbf{r}, \\
& \text{Feasibility cuts}, \\
& \text{Optimality cuts}.
\end{aligned}$$

The problem can be solved by adding feasibility and optimality cuts iteratively. Notice here in the subproblem (SUB-SUC), we have a bilinear term $w^T\mathbf{H}^T\mu$. In the following part, we propose two separation approaches to address the bilinear term.

5.3.1 Exact Separation Approach

Since the supporting space Ω of the random electricity load is described as a polytope as shown in (5–25), it can be verified that the optimal solution w^* to the subproblem (SUB-SUC) should satisfy the following proposition.

Proposition 5.2. *There exists an optimal solution* $(w^*, \lambda^*, \mu^*, \nu^*)$ *to the subproblem (SUB-SUC) such that* $w^*_{tb} = D^{b-}_t$ *or* $w^*_{tb} = D^{b+}_t$ *for each bus* b *in each time period* t.

Proof. For a fixed solution $(\lambda^*, \mu^*, \nu^*)$, the problem is to maximize a concave function with a polyhedral feasible region Ω, so at least one extreme point w^* of Ω is the optimal solution to the subproblem, i.e., $w^*_{tb} = D^{b-}_t$ or $w^*_{tb} = D^{b+}_t$. □

We let binary variable $z^{b+}_t = 1$ if $w^*_{tb} = D^{b+}_t$ and binary variable $z^{b-}_t = 1$ if $w^*_{tb} = D^{b-}_t$, and then based on Proposition 5.2, we have the following constraints hold:

$$z^{b+}_t + z^{b-}_t = 1, \tag{5–44}$$

$$w_{tb} = D^{b+}_t z^{b+}_t + D^{b-}_t z^{b-}_t, \tag{5–45}$$

where constraint (5–44) indicates the optimal load will either achieve its lower bound or upper bound. With constraints (5–44) and (5–45), we now address the bilinear term $w^T \mathbf{H}^T \mu$. Let $\mathbf{H}^T \mu = \kappa$, we have

$$\begin{aligned} w^T \mathbf{H}^T \mu &= \sum_{t\in T}\sum_{b\in B}(D^{b+}_t z^{b+}_t + D^{b-}_t z^{b-}_t)\kappa^b_t \\ &= \sum_{t\in T}\sum_{b\in B}(D^{b+}_t z^{b+}_t \kappa^b_t + D^{b-}_t z^{b-}_t \kappa^b_t) \\ &= \sum_{t\in T}\sum_{b\in B}(D^{b+}_t \kappa^{b+}_t + D^{b-}_t \kappa^{b-}_t) \qquad (5–46) \end{aligned}$$

$$s.t. \quad \kappa^b_t = (\mathbf{H}^T \mu)^b_t, \ \forall t \in T, \forall b \in B \tag{5–47}$$

$$\kappa^{b+}_t \geq -M z^{b+}_t, \ \forall t \in T, \forall b \in B \tag{5–48}$$

$$\kappa^{b+}_t \geq \kappa^b_t - M(1 - z^{b+}_t), \ \forall t \in T, \forall b \in B \tag{5–49}$$

$$\kappa^{b-}_t \geq -M z^{b-}_t, \forall t \in T, \ \forall b \in B \tag{5–50}$$

$$\kappa^{b-}_t \geq -\kappa^b_t - M(1 - z^{b-}_t), \ \forall t \in T, \forall b \in B. \tag{5–51}$$

Now we can replace the bilinear term $w^T \mathbf{H}^T \mu$ with (5–46) and add constraints (5–47) to (5–51) to the subproblem.

Next, we investigate how to generate Benders' feasibility cuts and optimality cuts. For the feasibility cuts, we only consider constraints (5–38) and (5–39) since only these constraints affect the feasibility. We use L-shape method to generate the feasibility cuts with the following feasibility check problem:

$$\begin{aligned} \varpi(y) = \max_{\hat{\lambda},\hat{\mu}} \ & (\mathbf{f}+\mathbf{F}y)^T\hat{\lambda} + \mathbf{g}\hat{\mu} + (D^+)^T\hat{\kappa}^+ + (D^-)^T\hat{\kappa}^- \\ s.t. \quad & \mathbf{D}^T\hat{\lambda} + \mathbf{G}^T\hat{\mu} \leq 0, \\ & \text{Constraints (5–47) to (5–51)}, \\ & \hat{\lambda}, \hat{\mu} \in [0,1]. \end{aligned}$$

If $\varpi(y) = 0$, y is a feasible solution. If $\varpi(y) > 0$, we can generate a feasibility cut $\varpi(y) \leq 0$.

For the optimality cuts, after we solve the master problem and obtain y and π^i for $i = 1, \cdots, N$, we substitute y into the subproblem and get $\phi^i(y)$. For each i, if $\phi^i(y) > \pi^i$, then y is not an optimal solution and we can generate a corresponding optimality cut $\phi^i(y) \leq \pi^i$.

5.3.2 Bilinear Separation Approach

In this subsection, we discuss the bilinear approach to generate Benders' feasibility cuts. Similarly, it is not necessary to consider constraint (5–40) since it does not affect the feasibility. Therefore, the feasibility check problem for these constraints is shown as follows (denoted as FEA):

$$\theta(y) = \max_{\hat{\lambda},\hat{\mu}} \ (\mathbf{f}+\mathbf{F}y)^T\hat{\lambda} + (\mathbf{g}+\mathbf{H}w)^T\hat{\mu} \tag{5–52}$$

$$s.t. \quad \mathbf{D}^T\hat{\lambda} + \mathbf{G}^T\hat{\mu} \leq 0, \tag{5–53}$$

$$\hat{\lambda}, \hat{\mu} \in [0,1]. \tag{5–54}$$

Next, we initiate the value of w as one of the extreme points, and with the fixed w, we solve (FEA) to obtain the optimal value $\theta(y)$ (denoted as $\theta^1(y, w)$) and optimal solution

$\hat{\lambda}^*$ and $\hat{\mu}^*$. Then, by fixing $\hat{\lambda} = \hat{\lambda}^*$ and $\hat{\mu} = \hat{\mu}^*$, we maximize the objective (5–52) with respect to $w \in \Omega$. In that case, we can obtain the optimal value (denoted as $\theta^2(y, \hat{\lambda}, \hat{\mu})$) and the optimal value w^*. If $\theta^2(y, \hat{\lambda}, \hat{\mu}) > \theta^1(y, w)$, let $w = w^*$ and process it iteratively. Otherwise, check whether $\theta^1(y, w) = 0$, if so, we can terminate the feasibility check; if not, add the feasibility cut $\theta^1(y, w) \leq 0$ to the master problem.

Then, we generate the Benders' optimality cuts by using the bilinear approach. Similarly, we initiate the value of w as one extreme point of Ω, and solve the problem (DR-SUC) to obtain the optimal value for each $i = 1, \cdots, N$ with the fixed w (denoted as $\phi_i^1(y, w)$) and optimal solution λ^*, μ^* and ν^*. Then, by fixing $\lambda = \lambda^*$, $\mu = \mu^*$ and $\nu = \nu^*$, we obtain the optimal value (denoted as $\phi_i^2(y, \lambda, \mu, \nu)$) and the optimal value w^* of the following problem

$$\phi_i^2(y, \lambda, \mu, \nu) = \max_{w \in \Omega} \mu^T \mathbf{H} w - \beta \rho^i(w) + (\mathbf{f} + \mathbf{F}y)^T \lambda + \mathbf{g}^T \mu + (\mathbf{h} + \mathbf{L}y)^T \nu.$$

If $\phi_i^2(y, \lambda, \mu, \nu) > \phi_i^1(y, w)$, let $w = w^*$ and process it iteratively until $\phi_i^2(y, \lambda, \mu, \nu) \leq \phi_i^1(y, w)$. Then check whether $\phi_i^2(y, \lambda, \mu, \nu) > \pi^i$, if so, generate the corresponding optimality cut $\phi_i^2(y, \lambda, \mu, \nu) \leq \pi^i$ to the master problem; otherwise, output the solutions.

5.4 Convergence Analysis

In this section, we examine the convergence properties of the DD-SUC to SUC as the number of historical data increases. We demonstrate that as the confidence set $\mathcal{D}$ shrinks with more observed historical data, the risk-averse problem DD-SUC converges to the risk-neutral problem SUC. We first analyze the convergence property of the second stage objective value.

Proposition 5.3. $\lim_{N\to\infty} \min_{\beta \geq 0} \left\{ \theta\beta + \frac{1}{N} \sum_{i=1}^{N} \max_{w \in \Omega} \{ \mathcal{Q}(y, u, v, w) - \beta \rho^i(w) \} \right\} = E_{\mathbb{P}_0}[\mathcal{Q}(y, u, v, w)]$*, that is,* $\lim_{\theta \to 0} \sup_{\mathbb{P} \in \mathcal{D}} E_{\mathbb{P}}[\mathcal{Q}(y, u, v, w)] = E_{\mathbb{P}_0}[\mathcal{Q}(y, u, v, w)]$.

Proof. First, since $\sup_{\mathbb{P} \in \mathcal{D}} E_{\mathbb{P}}[\mathcal{Q}(y, u, v, w)]$ is equivalent to (5–34), we know that (5–34) exists for any N. On the other side, according to the Law of Large Numbers

and Helly-Bray theorem, we have

$$\lim_{N\to\infty}\left\{\theta\beta+\frac{1}{N}\sum_{i=1}^{N}\max_{w\in\Omega}\{\mathcal{Q}(y,u,v,w)-\beta\rho(w,\xi^i)\}\right\}=E_{\mathbb{P}_0}[\max_{w\in\Omega}\{\mathcal{Q}(y,u,v,w)-\beta\rho(w,\xi)\}].$$

Therefore,

$$\begin{aligned}&\lim_{N\to\infty}\min_{\beta\geq 0}\left\{\theta\beta+\frac{1}{N}\sum_{i=1}^{N}\max_{w\in\Omega}\{\mathcal{Q}(y,u,v,w)-\beta\rho(w,\xi^i)\}\right\}\\&\leq\min_{\beta\geq 0}\lim_{N\to\infty}\left\{\theta\beta+\frac{1}{N}\sum_{i=1}^{N}\max_{w\in\Omega}\{\mathcal{Q}(y,u,v,w)-\beta\rho(w,\xi^i)\}\right\}\qquad(5\text{–}55)\\&=\min_{\beta\geq 0}E_{\mathbb{P}_0}[\max_{w\in\Omega}\{\mathcal{Q}(y,u,v,w)-\beta\rho(w,\xi)\}]\qquad(5\text{–}56)\end{aligned}$$

Next we show that for a given y, u, v,

$$\min_{\beta\geq 0}E_{\mathbb{P}_0}[\max_{w\in\Omega}\{\mathcal{Q}(y,u,v,w)-\beta\rho(w,\xi)\}]=E_{\mathbb{P}_0}[\mathcal{Q}(y,u,v,\xi)].\qquad(5\text{–}57)$$

For a given ξ, we assume the optimal solution to $\max_{w\in\Omega}\{\mathcal{Q}(y,u,v,w)-\beta\rho(w,\xi)\}$ is $w^*(\xi)$, then

$$E_{\mathbb{P}_0}[\max_{w\in\Omega}\{\mathcal{Q}(y,u,v,w)-\beta\rho(w,\xi)\}]=E_{\mathbb{P}_0}[\mathcal{Q}(y,u,v,w^*(\xi))]-\beta E_{\mathbb{P}_0}[\rho(w^*(\xi),\xi)\}].$$

Since for any $w\in\Omega$, $\mathcal{Q}(y,u,v,w^*(\xi))$ is bounded, we denote the upper bound of $\mathcal{Q}(y,u,v,w^*(\xi))$ as M, then

$$E_{\mathbb{P}_0}[\mathcal{Q}(y,u,v,w^*(\xi))]-\beta E_{\mathbb{P}_0}[\rho(w^*(\xi),\xi)\}]\leq M-\beta E_{\mathbb{P}_0}[\rho(w^*(\xi),\xi)\}].$$

First, we argue $E_{\mathbb{P}_0}[\rho(w^*(\xi),\xi)\}]=0$. If not, we can let β to be positive infinity, then $\min_{\beta\geq 0}E_{\mathbb{P}_0}[\max_{w\in\Omega}\{\mathcal{Q}(y,u,v,w)-\beta\rho(w,\xi)\}]$ will be unbounded. Therefore, $E_{\mathbb{P}_0}[\rho(w^*(\xi),\xi)\}]=0$. Since for any ξ, $\rho(w^*(\xi),\xi)\geq 0$, then we have $\rho(w^*(\xi),\xi)=0$, for any $\xi\in\Omega$. It means $w^*(\xi)=\xi$, $\forall\xi$. In that case,

$$\begin{aligned}&\min_{\beta\geq 0}E_{\mathbb{P}_0}[\max_{w\in\Omega}\{\mathcal{Q}(y,u,v,w)-\beta\rho(w,\xi)\}]\\&=\min_{\beta\geq 0}E_{\mathbb{P}_0}[\{\mathcal{Q}(y,u,v,w^*(\xi))-\beta\rho(w^*(\xi),\xi)\}]\end{aligned}$$

$$= \min_{\beta \geq 0} E_{\mathbb{P}_0}[\{\mathcal{Q}(y, u, v, \xi) - \beta\rho(\xi, \xi)\}]$$
$$= E_{\mathbb{P}_0}[\mathcal{Q}(y, u, v, \xi)].$$

From (5–56) and (5–57), we have

$$\lim_{\theta \to 0} \sup_{\mathbb{P} \in \mathcal{D}} E_{\mathbb{P}}[\mathcal{Q}(y, u, v, w)] \leq E_{\mathbb{P}_0}[\mathcal{Q}(y, u, v, w)].$$

On the other hand, since $\mathbb{P}_0 \in \mathcal{D}$, we have

$$\sup_{\mathbb{P} \in \mathcal{D}} E_{\mathbb{P}}[\mathcal{Q}(y, u, v, w)] \geq E_{\mathbb{P}_0}[\mathcal{Q}(y, u, v, w)],$$

which immediately yields

$$\lim_{\theta \to 0} \sup_{\mathbb{P} \in \mathcal{D}} E_{\mathbb{P}}[\mathcal{Q}(y, u, v, w)] \geq E_{\mathbb{P}_0}[\mathcal{Q}(y, u, v, w)].$$

Therefore, we have the theorem holds. □

We denote the optimal value to problem (DD-SUC) with N number of historical data as $\psi(N)$ and the optimal value to problem (SUC) as $\psi(0)$, we have the following theorem:

Proposition 5.4. $\lim_{N \to \infty} \psi(N) = \psi(0)$.

Proof. First, notice that $N \to \infty$ is equivalent to $\theta \to 0$. Therefore, to prove $\lim_{N \to \infty} \psi(N) = \psi(0)$ is equivalent to prove $\lim_{\theta \to 0} \psi(\theta) = \psi(0)$, where $\psi(\theta)$ is the same as $\psi(N)$. Denote $V_\theta(y, u, v)$ as the objective value of problem (DD-SUC) and $W(y, u, v)$ as the objective value of problem (SUC), with respect to any given solution y,u,v. According to Proposition 5.3, for any arbitrary small number ϵ, and any given solution y, u, v, there exists $\Delta_\theta > 0$, such that

$$|V_\theta(y, u, v) - W(y, u, v)| \leq \epsilon, \quad \forall \theta \leq \Delta_\theta. \tag{5–58}$$

Then for any $\theta \leq \Delta_\theta$, denote the optimal solution to (DD-SUC) as y_θ^*, u_θ^*, v_θ^*, and the optimal solution to (SUC) as $\hat{y}$, $\hat{u}$, $\hat{v}$, then

$$\begin{aligned}
|\psi(\theta) - \psi(0)| &= \psi(\theta) - \psi(0) \\
&= V_\theta(y_\theta^*, u_\theta^*, v_\theta^*) - W(\hat{y}, \hat{u}, \hat{v}) \\
&\leq V_\theta(\hat{y}, \hat{u}, \hat{v}) - W(\hat{y}, \hat{u}, \hat{v}) \\
&\leq |V_\theta(\hat{y}, \hat{u}, \hat{v}) - W(\hat{y}, \hat{u}, \hat{v})| \\
&\leq \epsilon
\end{aligned}$$

Therefore, we can prove the proposition. □

5.5 Case Study

In this section, we test the system performance with a modified IEEE 118-bus system, based on the one given online at *http://motor.ece.iit.edu/data*. The system contains 118 buses, 33 generators, and 186 transmission lines. The operational time interval is 24 hours. In our experiments, we set the feasibility tolerance gap to be 10^{-6} and the optimality tolerance gap to be 10^{-4}. The MIP gap tolerance for the master problem is the CPLEX default gap. We use C++ with CPLEX 12.1 to implement the proposed formulations and algorithms. All experiments are executed on a computer workstation with 4 Intel Cores and 8GB RAM. In our experiment, we compare the performances of our proposed approach with the stochastic optimization approach.

We introduce a penalty cost with the value $\$5,000$/MWh [7], for any power imbalance or transmission capacity/ramp-rate limit violation. Besides, we assume the uncertain load follows a multivariate normal distribution, and we generate samples as the observed historical data. In addition, we set the confidence level as 0.99. Then we compare our proposed model with the traditional two-stage stochastic unit commitment model. We obtain the total costs corresponding to our proposed data-driven risk-averse stochastic optimization approach (DD-SUC) and the traditional two-stage stochastic

optimization approach (SO) through the following steps: 1) Obtain the unit commitment decisions by using the DD-SUC approach and SO approach respectively. 2) Fix the obtained UC decisions and solve the second-stage problem repeatedly for 50 randomly generated samples to obtain the total costs for each approach respectively.

We report the results in Table 5-1. The first column represents the number of samples. The third column computes the value of θ based on Proposition 5.1. The numbers of start-ups are given in the fourth column and unit commitment costs are given in the fifth column. Finally, the sixth column gives the total costs.

Table 5-1. Comparison between SO and DD-SUC approaches

# of samples	Model	θ	# of start-ups	UC.C.($)	T.C.($)
1	SO	26.626	21	5440	602828
	DD-SUC	26.626	29	5905	604425
10	SO	8.420	21	5440	602823
	DD-SUC	8.420	29	5865	544465
50	SO	3.765	22	5495	602884
	DD-SUC	3.765	28	5805	544394
100	SO	2.663	23	5555	602960
	DD-SUC	2.663	28	5805	543703

From Table 3-2, we have the following observations:

(1) On one hand, as the number of samples increases, the number of start-ups obtained by the proposed DD-SUC approach decreases, and the unit commitment cost decreases, as well as the total cost. That is because more historical data are observed, so that the uncertainty set shrinks, which leads to less conservative solutions. On the other hand, as the number of samples increases, the number of start-ups obtained by the SO approach increases, and the unit commitment cost increases, since more historical data leads to more robust solutions for the SO approach.

(2) Second, for the tested number of samples, the UC decisions derived by the SO approach are not feasible to some simulated load scenarios. However, the UC decisions obtained from the proposed approach have no feasibility issue. Therefore, the SO approach has more total costs than the proposed DD-SUC approach, since the penalty costs occur by using the SO approach.

(2) Third, as compared to the SO approach, the DD-SUC approach brings more generators online to provide sufficient generation capacity to maintain the system balance. As a result, the DD-SUC approach has a larger UC cost than the SO approach. This result verifies that the proposed DD-SUC approach is more robust than the SO approach.

5.6 Summary

In this chapter, we propose a data-driven risk-averse two-stage stochastic optimization approach to cope with the unit commitment problem. By learning from the historical data that are drawn from an ambiguous probability distribution of the load and renewable energy output, we construct a confidence set of the probability distribution with Wasserstein metric and develop a data-driven risk-averse two-stage stochastic unit commitment model. In addition, we reformulate the model into a traditional two-stage robust optimization problem and demonstrate that as the number of historical data increases, the data-driven risk-averse two-stage stochastic unit commitment model converges to the traditional two-stage stochastic unit commitment model. Finally, a revised IEEE 118-bus system is examined to show the efficiency of the proposed model.

CHAPTER 6
CONCLUSIONS

In this dissertation, several optimization-under-uncertainty formulations are proposed to capture the uncertainty in the unit commitment problem. First, we propose a robust approach to cope with two dimensional uncertainties, i.e., electricity load and wind power output. Then, we develop a unified stochastic and robust unit commitment problem, which can take advantage of both robust optimization approach and stochastic optimization approach and meanwhile overcome their disadvantages. Then, we extend our research to the fundamental data-driven risk-averse two-stage stochastic optimization, for which the distribution of the random variable varies within a given confidence set. Based on the historical data, we construct the confidence set, with a new class of probability metrics. In addition, we provide the solution methodology to deal with the problem for both discrete and continuous distribution cases. Moreover, we prove that our risk-averse stochastic program converges to the risk-neutral case as the size of historical data increases to infinity. Finally, we apply the proposed data-driven risk-averse two-stage stochastic optimization model to power system operational problems. Possible future directions include: 1) extending current research to the chance-constrained and multi-stage cases, and analyzing their applications on power systems operations and supply chain management. 2) developing exact approaches to solve large-sized problems in power systems.

REFERENCES

[1] Adler, R. J. and Brown, L. D. "Tail behaviour for suprema of empirical processes." *The Annals of Probability* 14 (1986).1: 1–30.

[2] Albadi, M. H. and El-Saadany, E. F. "Demand response in electricity markets: An overview." *IEEE Power Engineering Society General Meeting*. Montreal, 2007, 1–5.

[3] Ambrosio, L., Gigli, N., and Savar, G. *Gradient flows in metric spaces and in the spaces of probability measures*. Springer, 2000.

[4] Baringo, L. and Conejo, A. "Offering strategy via robust optimization." *IEEE Transactions on Power Systems* 26 (2011).3: 1418–1425.

[5] Barth, R., Brand, H., Meibom, P., and Weber, C. "A stochastic unit-commitment model for the evaluation of the impacts of integration of large amounts of intermittent wind power." *International Conference on Probabilistic Methods Applied to Power Systems*. Stockholm, Sweden, 2006.

[6] Ben-Ta, A., Hertog, D. Den, Waegenaere, A. De, Melenberg, B., and Rennen, G. "Robust solutions of optimization problems affected by uncertain probabilities." *Management Science* 59 (2013).2: 341–357.

[7] Bertsimas, D., Litvinov, E., Sun, X. A., Zhao, J., and Zheng, T. "Adaptive robust optimization for the security constrained unit commitment problem." *IEEE Transactions on Power Systems* 28 (2013).1: 52–63.

[8] Bertsimas, D. and Sim, M. "The price of robustness." *Operations Research* 52 (2004).1: 35–53.

[9] Birge, J. and Louveaux, F. *Introduction to Stochastic Programming*. Springer, 1997.

[10] Bolley, F., Guillin, A., and Villani, C. "Quantitative concentration inequalities for empirical measures on non-compact spaces." *Probability Theory and Related Fields* 137 (2007).3-4: 541–593.

[11] Bolley, F. and Villani, C. "Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities." *Annales de la Faculté des Sciences de Toulouse*. vol. 14. 2005, 331–352.

[12] Bouffard, F. and Galiana, F. D. "Stochastic security for operations planning with significant wind power Generation." *IEEE Transactions on Power Systems* 23 (2008).2: 306–316.

[13] Bouffard, F., Galiana, F. D., and Conejo, A. J. "Market-clearing with stochastic security-Part I: Formulation." *IEEE Transactions on Power Systems* 20 (2005).4: 1818–1826.

[14] ———. "Market-clearing with stochastic security-Part II: Case studies." *IEEE Transactions on Power Systems* 20 (2005).4: 1827–1835.

[15] Chen, Y., Wang, X., and Guan, Y. "Applying robust optimization to MISO look-ahead unit commitment." *Proceedings of FERC Technical Conference*. 2013.

[16] Cover, T. M. and Thomas, J. A. *Elements of information theory*. John Wiley & Sons, 2012.

[17] Cybakov, A. B. *Introduction to nonparametric estimation*. Springer, 2009.

[18] Delage, E. and Ye, Y. "Distributionally robust optimization under moment uncertainty with application to data-driven problems." *Operations Research* 58 (2010).3: 595–612.

[19] Dembo, A. and Ofer, Z. *Large deviations techniques and applications*, vol. 38. Springer, 2010.

[20] Deng, Y. and Du, W. "The Kantorovich metric in computer science: A brief survey." *Electronic Notes in Theoretical Computer Science* 253 (2009).3: 73–82.

[21] Devroye, L. and Gyorfi, L. "No empirical probability measure can converge in the total variation sense for all distributions." *The Annals of Statistics* 18 (1990).3: 1496–1499.

[22] Dudley, R. M. *Real analysis and probability*, vol. 74. Cambridge University Press, 2002.

[23] Fink, S., Mudd, C., Porter, K., and Morgenstern, B. "Wind energy curtailment case studies." *NREL Report* (2009): SR–550–46716.

[24] Gibbs, A. L. and Su, F. E. "On choosing and bounding probability metrics." *International statistical review* 70 (2002).3: 419–435.

[25] Goh, J. and Sim, M. "Distributionally robust optimization and its tractable approximations." *Operations Research* 58 (2010).4-Part-1: 902–917.

[26] Gray, R. M., Neuhoff, D. L., and Shields, P. C. "A generalization of Ornstein's $\bar{d}$ distance with applications to information theory." *The Annals of Probability* 3 (1975).2: 315–328.

[27] Guan, X., Luh, P. B., Yan, J. A., and Amalfi, J. A. "An optimization-based method for unit commitment." *International Journal of Electrical Power & Energy Systems* 14 (1992).1: 9–17.

[28] Hajimiragha, A., Cañizares, C., Fowler, M., Moazeni, S., and Elkamel, A. "A robust optimization approach for planning the transition to plug-in hybrid electric vehicles." *IEEE Transactions on Power Systems* 26 (2011).4: 2264–2274.

[29] Jiang, R. and Guan, Y. "Data-driven chance constrained stochastic program." *Technical Report, Available at Optimization-Online* (2013).

[30] Jiang, R., Wang, J., and Guan, Y. "Robust unit commitment with wind power and pumped storage hydro." *IEEE Transactions on Power Systems* 27 (2012).2: 800–810.

[31] ———. "Robust unit commitment with wind power and pumped storage hydro." *IEEE Transactions on Power Systems* 27 (2012): 800–810.

[32] Jiang, R., Zhang, M., Li, G., and Guan, Y. "Two-stage robust power grid optimization problem." *Technical Report, Available at Optimization-Online* (2010).

[33] Kantorovich, L. V. and Rubinshtein, G. S. "On a space of totally additive functions." *Vestn Lening. Univ.* 13 (1958).7: 52–59.

[34] Khodaei, A., Shahidehpour, M., and Bahramirad, S. "SCUC with hourly demand response considering intertemporal load characteristics." *IEEE Transactions on Smart Grid* 2 (2011).3: 564–571.

[35] Kirschen, D. S. "Demand-side view of electricity markets." *IEEE Transactions on Power Systems* 18 (2003).2: 520–527.

[36] Kirschen, D. S., Strbac, G., Cumperayot, P., and de Paiva Mendes, D. "Factoring the elasticity of demand in electricity prices." *IEEE Transactions on Power Systems* 15 (2000).2: 612–617.

[37] Lindvall, T. *Lectures on the coupling method*. Courier Dover Publications, 2002.

[38] Love, D. and Bayraksan, G. "Phi-divergence constrained ambiguous stochastic programs." *Optimization Online* (2013).

[39] California Public Utilities Commission. "Decision setting procurement quantity requirements for retail sellers for the renewables portfolio standard program." (2011).

URL http://docs.cpuc.ca.gov/WORD_PDF/FINAL_DECISION/154695.PDF

[40] Federal Energy Regulatory Commission. "Wholesale competition in regions with organized electric markets: FERC's advanced notice of proposed rulemaking." (2007).

URL http://www.kirkland.com/siteFiles/Publications/C430B16C519842DE1AEB2623F7DE21D6.pdf.

[41] ———. "A national assessment of demand response potential." (2009).

URL www.ferc.gov/legal/staff-reports/06-09-demand-response.pdf.

[42] MidwestISO. "Business practices manual 003: energy market instruments.", 2008.

URL https://www.midwestiso.org.

[43] U.S. Department of Energy. "Benefits of demand response in electricity markets and recommendations for achieving them." (2006).

URL http://eetd.lbl.gov/ea/ems/reports/congress-1252d.pdf.

[44] ———. "20% wind energy by 2030: Increasing wind energy's contribution to U.S. electricity supply." (2008).

URL http://eere.energy.gov/wind/pdfs/41869.pdf.

[45] Meyn, S. S. P. and Tweedie, R. L. *Markov chains and stochastic stability*. Cambridge University Press, 2009.

[46] Moreno, P. J., Ho, P. P., and Vasconcelos, N. "A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications." *Advances in neural information processing systems*. 2003.

[47] Navid-Azarbaijani, N. *Load model and control of residential appliances*. McGill University, 1996.

[48] Papavasiliou, A. *Coupling renewable energy supply with deferrable demand*. Ph.D. thesis, University of California, Berkeley, 2012.

[49] Papavasiliou, A., Oren, S. S., and O'Neill, R. P. "Reserve requirements for wind power integration: A scenario-based stochastic programming framework." *IEEE Transactions on Power Systems* 26 (2011).4: 2197–2206.

[50] Pardo, L. *Statistical inference based on divergence measures*. CRC Press, 2005.

[51] Parzen, E. "On estimation of a probability density function and mode." *The Annals of Mathematical Statistics* 33 (1962).3: 1065–1076.

[52] Rachev, S. T. *Mass transportation problems*, vol. 2. Springer, 1998.

[53] Rockafellar, R. T. and Wets, R. J. "Scenarios and policy aggregation in optimization under uncertainty." *Mathematics of Operations Research* 16 (1991).1: 119–147.

[54] Rosenblatt, M. "Remarks on some nonparametric estimates of a density function." *The Annals of Mathematical Statistics* 27 (1956).3: 832–837.

[55] Ruiz, P. A., Philbrick, C. R., Zak, E., Cheung, K. W., and Sauer, P. W. "Uncertainty management in the unit commitment problem." *IEEE Transactions on Power Systems* 24 (2009).2: 642–651.

[56] Shorack, G. R. *Probability for statisticians*. Springer New York, 2000.

[57] Shorack, G. R. and Wellner, J. A. *Empirical processes with applications to statistics*, vol. 59. SIAM, 2009.

[58] So, A. M. C., Zhang, J., and Ye, Y. "Stochastic combinatorial optimization with controllable risk aversion level." *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2006. 224–235.

[59] Street, A., Oliveira, F., and Arroyo, J. "Contingency-constrained unit commitment with n-K security criterion: A robust optimization approach." *IEEE Transactions on Power Systems* 26 (2011): 1581–1590.

[60] Su, C. L. and Kirschen, D. "Quantifying the effect of demand response on electricity markets." *IEEE Transactions on Power Systems* 24 (2009).3: 1199–1207.

[61] Thimmapuram, P. R., Kim, J., Botterud, A., and Nam, Y. "Modeling and simulation of price elasticity of demand using an agent-based model." *Innovative Smart Grid Technologies (ISGT), 2010*. IEEE, 2010, 1–8.

[62] Thollard, F., Dupont, P., and Higuera, C. D. L. "Probabilistic DFA inference using Kullback-Leibler divergence and minimality." *Proceedings of the Seventeenth International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., 2000, 975–982.

[63] Tuohy, A., Denny, E., and O'Malley, M. "Rolling unit commitment for systems with significant installed wind capacity." *IEEE Lausanne Power Tech*. 2007, 1380–1385.

[64] Tuohy, A., Meibom, P., Denny, E., and O'Malley, M. "Unit commitment for systems with significant wind penetration." *IEEE Transactions on Power Systems* 24 (2009).2: 592–601.

[65] Ummels, B. C., Gibescu, M., Pelgrum, E., Kling, W. L., and Brand, A. J. "Impacts of wind power on thermal generation unit commitment and dispatch." *IEEE Transactions on Energy Conversion* 22 (2007).1: 44–51.

[66] Van der Vaart, A. W. *Asymptotic statistics*, vol. 3. Cambridge university press, 2000.

[67] Vershik, A. M. "Kantorovich metric: Initial history and little-known applications." *Journal of Mathematical Sciences* 133 (2006).4: 1410–1417.

[68] Villani, C. *Topics in optimal transportation*. 58. American Mathematical Society, 2003.

[69] Wang, J., Botterud, A., Miranda, V., Monteiro, C., and Sheble, G. "Impact of wind power forecasting on unit commitment and dispatch." *8th Int. Workshop on Large-Scale Integration of Wind Power into Power Systems*. Bremen, Germany, 2009.

[70] Wang, J., Shahidehpour, M., and Li, Z. "Security-constrained unit commitment with volatile wind power generation." *IEEE Transactions on Power Systems* 23 (2008).3: 1319–1327.

[71] Wang, Q., Guan, Y., and Wang, J. "A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output." *IEEE Transactions on Power Systems* 27 (2012): 206–215.

[72] Wang, Q., Wang, J., and Guan, Y. "Stochastic unit commitment with uncertain demand response." *IEEE Transactions on Power Systems* 28 (2013).1: 562–563.

[73] Wang, S. J., Shahidehpour, S. M., Kirschen, D. S., Mokhtari, S., and Irisarri, G. D. "Short-term generation scheduling with transmission and environmental constraints using an augmented Lagrangian relaxation." *IEEE Transactions on Power Systems* 10 (1995).3: 1294–1301.

[74] Wolfowitz, J. "Generalization of the theorem of Glivenko-Cantelli." *The Annals of Mathematical Statistics* (1954): 131–138.

[75] Wu, L., Shahidehpour, M., and Li, T. "Stochastic security-constrained unit commitment." *IEEE Transactions on Power Systems* 22 (2007).2: 800–811.

[76] Zhao, C. and Guan, Y. "Unified stochastic and robust unit commitment." *IEEE Transactions on Power Systems* 28 (2013).3: 3353 – 3361.

[77] Zhao, C., Wang, J., Watson, J. P., and Guan, Y. "Multi-stage robust unit commitment considering wind and demand response uncertainties." *IEEE Transactions on Power Systems* 28 (2013).3: 2708 – 2717.

[78] Zhao, L. and Zeng, B. "Robust unit commitment problem with demand response and wind energy." *Technical Report, Available at Optimization-Online* (2010).

[79] Zolotarev, V. M. and Vladimir, M. "Probability metrics." *Teoriya Veroyatnostei i ee Primeneniya* 28 (1983).2: 264–287.

BIOGRAPHICAL SKETCH

Ms. Chaoyue Zhao received her Ph.D. from the University of Florida in summer 2014 in Industrial and Systems Engineering. Before that, she obtained her B.S. degree in Information and Computing Sciences from the Fudan University, China, in 2010. Her research interests include data-driven stochastic optimization and stochastic integer program with their applications on smart grid, energy systems, and supply chain management.

She has collaborated with Sandia and Argonne National Labs for her research and worked at Pacific Gas & Electric Company, one of the largest utilities in US. Her research has led to papers published on the flagship journal for power systems and a best student paper award in the CIS track of ISERC conference 2012. She was awarded Graduate Student Council Outstanding Research Award in 2014, and was also one of the recipients of the Office of Research Travel Grants, the Graduate Student Council Travel Grants, NSF Student Travel Award, and the 2013 Mixed Integer Programming Workshop Student Travel Award.