Pontifícia Universidade Católica
do Rio de Janeiro

Carlos Andrés Gamboa Rodríguez

# Partition-based Method for Two-stage Stochastic Linear Programming Problems with Complete Recourse



**Dissertação de Mestrado**

Dissertation presented to the Programa de Pós–Graduação em Engenharia de Produção of the Departmento Engenharia Industrial in partial fulfillment of the requirements for the degree of Mestre em Engenharia de Produção.

Advisor: Prof. Davi Michel Valladão

Rio de Janeiro
April 2017

Pontifícia Universidade Católica
do Rio de Janeiro

**Carlos Andrés Gamboa Rodríguez**

# Partition-based Method for Two-stage Stochastic Linear Programming Problems with Complete Recourse

Dissertation presented to the Programa de Pós–Graduação em Engenharia de Produção of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Engenharia de Produção. Approved by the undersigned Examination Committee.

**Prof. Davi Michel Valladão**
Advisor
Departamento de Engenharia Industrial — PUC–Rio

**Prof. Bruno da Costa Flach**
IBM Research Brasil

**Prof. Alexandre Street de Aguiar**
Departamento de Engenharia Elétrica — PUC-Rio

**Prof. Márcio da Silveira Carvalho**
Vice Dean of Graduate Studies
Centro Técnico Científico — PUC-Rio

Rio de Janeiro, April 25th, 2017

**Carlos Andrés Gamboa Rodríguez**

Carlos Andrés currently is a master candidate of Industrial Engineering Department at PUC-Rio. His research focus on robust optimization techniques and stochastic programming. He graduated in mathematics at National University of Colombia in 2013. After his undergraduate he worked as researcher in the Research Department of Latin American Reserve Fund. His work in that financial institution was about optimization model update for risk allocation budget of the portfolios, studies of the intensive models for default risk, GARCH (Generalized Autoregressive Conditional Heteroskedasticity) methodologies to obtain dynamics correlation matrices for the risk factors that define the different invesment strategies, among others research activities.



Bibliographic data

Gamboa, Carlos Andrés

Partition-based Method for Two-stage Stochastic Linear Programming Problems with Complete Recourse / Carlos Andrés Gamboa Rodríguez; advisor: Davi Michel Valladão. — 2017.

79 f. : il. (color.); 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Departamento de Engenharia Industrial, 2017.

Inclui bibliografia.

1. Engenharia Industrial – Teses. 2. Otimização Robusta. 3. Método de Cenários (SAA). 4. Programação Estocástica. 5. Método de Partição. I. Valladão, Davi Michel. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. III. Título.

CDD: 004

# Acknowledgement



Thank to my advisors Prof. Davi Michel Valladão for their support and encouragement for this work.

Thank to CAPES, and PUC-Rio for the financial support which allowed this work to be done.

To my parents, family and friends who supported me even with my absence in family life.

To all colleagues, faculty and staff of the Department of PUC-Rio, for the fellowship, learning and support.

## Abstract

Gamboa, Carlos Andrés; Valladão, Davi Michel (Advisor). **Partition-based Method for Two-stage Stochastic Linear Programming Problems with Complete Recourse**. Rio de Janeiro, 2017. 79p. Dissertação de Mestrado — Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

The hardest part of modelling decision-making problems in the real world, is the uncertainty associated to realizations of futures events. The stochastic programming is responsible about this subject; the target is finding solutions that are feasible for all possible realizations of the unknown data, optimizing the expected value of some functions of decision variables and random variables.

The approach most studied is based on Monte Carlo simulation and the Sample Average Approximation (SAA) method which is a kind of discretization of expected value, considering a finite set of realizations or scenarios uniformly distributed.

It is possible to prove that the optimal value and the optimal solution of the SAA problem converge to their counterparts of the true problem when the number of scenarios is sufficiently big.

Although that approach is useful, there exist limiting factors about the computational cost to increase the scenarios number to obtain a better solution; but the most important fact is that SAA problem is function of each sample generated, and for that reason is random, which means that the solution is also uncertain, and to measure its uncertainty it is necessary consider the replications of SAA problem to estimate the dispersion of the estimated solution, increasing even more the computational cost.

The purpose of this work is presenting an alternative approach based on robust optimization techniques and applications of Jensen's inequality, to obtain bounds for the optimal solution, partitioning the support of distribution (without scenarios creation) of unknown data, and taking advantage of the convexity. At the end of this work the convergence of the bounding problem and the proposed solution algorithms are analyzed.

## Keywords

Robust Optimization; Sample Average Approximation; Stochastic Programming; Partition-based method.

# Resumo

Gamboa, Carlos Andrés; Valladão, Davi Michel. **Método de Partição para Problemas de Programação Linear Estocástica dois Estágios com Recurso Completo**. Rio de Janeiro, 2017. 79p. Dissertação de Mestrado — Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

A parte mais difícil de modelar os problemas de tomada de decisão do mundo real, é a incerteza associada à realização de eventos futuros. A programação estocástica se encarrega desse assunto; o objetivo é achar soluções que sejam factíveis para todas as possíveis realizações dos dados, otimizando o valor esperado de algumas funções das variáveis de decisão e de incerteza. A abordagem mais estudada está baseada em simulação de Monte Carlo e o método SAA (*Sample Average Approximation*) o qual é uma formulação do problema verdadeiro para cada realização da data incerta, que pertence a um conjunto finito de cenários uniformemente distribuídos.
É possível provar que o valor ótimo e a solução ótima do problema SAA converge a seus homólogos do problema verdadeiro quando o número de cenários é suficientemente grande.
Embora essa abordagem seja útil ali existem fatores limitantes sobre o custo computacional para obter soluções mais precisas aumentando o número de cenários; no entanto o fato mais importante é que o problema SAA é função de cada amostra gerada e por essa razão é aleatório, o qual significa que a sua solução também é incerta, e para medir essa incerteza é necessário considerar o número de replicações do problema SAA afim de estimar a dispersão da solução, aumentando assim o custo computacional.
O propósito deste trabalho é apresentar uma abordagem alternativa baseada em um método de partição que permite obter cotas para estimar deterministicamente a solução do problema original, com aplicação da desigualdade de Jensen e de técnicas de otimização robusta. No final se analisa a convergência dos algoritmos de solução propostos.



# Palavras–chave

Otimização Robusta; Método de Cenários (SAA); Programação Estocástica; Método de Partição.

# Contents

1 Introduction **10**

2 Modelling the Uncertainty **12**
2.1 Inventory Model 12
2.2 Two-Stage Stochastic Linear Model 14
2.3 Formulation Problem to Discrete Distributions (a Motivator Case) 15
2.4 Scenario Model and Sample Average Approximation (SAA) 16
2.5 Statistical Properties of SAA Estimators 17
2.6 The Farmer's Problem 21

3 The Partition-based Method **24**
3.1 Notation and Preliminary Definitions 24
3.2 Motivation 25
3.3 Method 28

4 Algorithms **37**
4.1 Solution Algorithm with a Radom Partition Refinement (SARPR) 37
4.2 Solution Algorithm with Worst-case Partition Refinement (SAWPR) 41
4.3 How to Choose the Dimension 41
4.4 How to Choose the Cell 44
4.5 Graphyc Explanation of the SAWPR 46

5 Case Study **53**
5.1 Bounding the Uncertainty 53
5.2 Upper Bound 54
5.3 Lower Bound 56

6 Computational Results **58**
6.1 SAA for the Farmer's Problem 58
6.2 Numerical Results from the SARPR 59
6.3 Numerical Results from the SAWPR 63

7 Final remarks **66**

8 Future Works **67**

Bibliography **67**

9 Bibliography **68**

A Mathematical and Statistical Background **70**
A.1 The Weak Law of Large Numbers 70
A.2 The Central Limit Theorem 70
A.3 Continuity of Convex Functions 72
A.4 Jensen's Inequality 74

A.5 The Total Probability Law 75
A.6 Large Linear Programming Problem for the Upper Bound 76
A.7 Average Optimizer 77
A.8 Edmundson-Madansky Inequality 78

# List of figures


3.1 Extreme points of the cell $I_k$ 36

4.1 Example of refinement $i^* = M$ 38
4.2 Partitioning of support with: (a) 2 cells, (b) 3 cells, (c) 4 cells, (d) 5 cells, (e) 6 cells, (f) 100 cells, (g) 1000 cells, (h) 15000 cells. 40
4.3 Convex hull (shaded region) for the set $V(I_0)$, where $I_0 = \Xi$ is the original support of $\xi$. The black point is the optimal solution for the maximization problem of $Q(\bar{x}_0^*, \xi)$ with $\xi \in$ conv$V(I_0)$. 47
4.4 Initial cut in the mean $\bar{\xi}_B = 20$. 48
4.5 (a) Convex hull (shaded region) for the set $V(I_1)$, where $I_1 = [2,3] \times [2.4, 3.6] \times [14, 20]$. The black point is the optimal solution for the maximization problem of $Q(\bar{x}^*, \xi)$ with $\xi \in$ conv$V(I_1)$. (b) Choose of the cell $I_1$, as the optimal cell. 49
4.6 Choose of the cell $I_1 = [2,3] \times [2.4, 3.6] \times [16, 18]$, as the optimal cell. 50
4.7 Partition of the support with: (a) 4 cells, (b) 20 cells, (c) 100 cells, (d) 300 cells, (e) 500 cells, (f) 900 cells. 52

6.1 SAA estimator for the Framer's Problem 58
6.2 Upper and lower bounds estimation. 60
6.3 Error graph. 61
6.4 Comparison of upper and lower bound convergence. 62
6.5 Bounding from SAWPR Algorithm. 64
6.6 Estimation error. 65

A.1 Convex function. 74

# 1
# Introduction

The two-stage stochastic programming problem is a model that represents the decision process under uncertainty, where in the first place we have to take decisions that do not depend on unknow information, and in the second one, we can wait to know the missing information to take the decisions that complete the process. This can be expressed mathematically by the following optimization problem:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Min}} \quad & c^t x + \mathbb{E}[Q(x,\xi)] \\ \text{s.t.} \quad & Ax = b, \quad x \geq 0, \end{aligned} \tag{1-1}$$

where

$$\begin{aligned} Q(x,\xi) = \underset{y\in\mathbb{R}^m}{\text{Min}} \quad & q^t y \\ \text{s.t.} \quad & Tx + Wy = h, \quad y \geq 0, \end{aligned} \tag{1-2}$$

is known as the recourse function and $\xi := (q, h, T, W)$ represents the unknown data or uncertainty.

The Sample Average Approximation (SAA) is a statistical method that considers the creation of scenarios to fit the real distribution of $\xi$ with a finite set of realizations. Nevertheless this is an useful technique and natural way to solve the problem, that kind of discretization of the expected operator demands a big scenarios number to converge with a high computational cost and offers a random estimated solution.

In this work we want to propose a partition-based method of the support of the distribution for a two-stage stochastic linear programming problem, to get a numerical approximation that converges, in a deterministic way, to the real solution, as alternative to Sample Average Approximation (SAA) method, that gives a random estimator for the true solution, without deterministic bounds.

This method is replicable since the sequential refinement problem of the partition can be reproduced to get the same result always, which is a very important property since this allows to adopt a unique policy of invesment according to the decision making problem modelled.

Nowadays there exist some numerical techniques that implement bounds approximations and partition refinements based on convex-concave properties and dual information of the original problem; those techniques are mencionated in classic works like (Birge and West, 1986), (Edirisinghe and Ziemba, 1996), and (Carøe and Tind, 1998). More recently in (Song and Luedtke, 2015) an adaptative partition-based method is considered as relaxation of the original problem to reduce the computational cost of the sampling method.

By other hand, our partition-based model consider the bounding problem that results of the Jensen's inequality application and the robust formulation of second stage problem (this last inspired by Bertsimas and Sim, 2004), exploring the convexity of the recourse function. According to this, the goal is finding the best way to discretize the support of uncertain variable to converge with the smallest partition size and get a high-quality solution without high computational cost.

This objective led to construct two algorithms: the Solution Algorithm with a Random Partition Refinement (SARPR) explore uniformly the distribution of the uncertain data without any additional information. The Solution Algorithm with Worst-case Partition Refinement (SAWPR) consider extreme realizations of the uncertain data to identify optimal cuts which adapts the robustness of the model and adjust the conservatism level for the estimated solution.

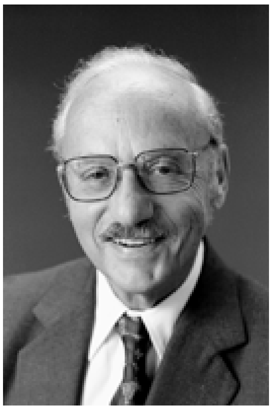
The algorithms were compared and the computational results allowed prove empirically the convergence of the proposed partition-based method.

The remainder of this dissertation is structured as follows. In **Chapter 2** the basic concepts about two-stage stochastic programming problems and properties of Sample Average Approximation Method are presented. **Chapter 3** the notation and definitions are introduced; the partition-based model is presented and there is a mathematical proof of its convergence. In **Chapter 4** the proposed algorithms are described. In **Chapter 5** a study case about the bounding problem formulation for the Farmer's problem is analyzed. In **Chapter 6** the computational results are presented and both proposed algorithms are compared. Finally, in **Chapter 7** and **Chapter 8** the final remarks and future works are discussed.

# 2
# Modelling the Uncertainty

**George Bernard Dantzig***(November 8, 1914 - May 13, 2005) American mathematician recognized by his constributions to operational research and mathematical programming. His work "Linear Programming under Uncertainty", established the foundations of the stochastic programming.*

In this chapter the basics concepts about uncertainty modelling are presented.



The stochastic programming is a theoretical framework to solve optimization problems with uncertain data for which their probability distribution is known or it can be estimated. In this section we will present some examples of decision-making problems under uncertainty, to give an intuition about their modelling and define some basic concepts related with that topic.

## 2.1
## Inventory Model

One company wants to order $x$ quantity of some product to attend the demand $d$. The product unitary cost is $c > 0$. If $d$ is bigger than $x$ the penalty cost per unit to the company is $b \geq 0$. So, the additional cost would be $b(d-x)$ if $d > x$, and zero in otherwise. For other hand, if the company ordered more products than the demand, the storage cost is $h$, and in that situation the incured cost woul be $h(x-d)$. Therefore the company total cost is (Shapiro et al., 2009)

$$F(x,d) = cx + b[d-x]_+ + h[x-d]_+ \tag{2-1}$$

where $[a]_+$ denote the maximum $max\{a, 0\}$. To make sense, we assume that $b > c$.

The target is minimizing the total cost considering $x$ as decision variable and $d$ as parameter. In that context, if the demand was known, the solution would be $\bar{x} = d$.

Now, suppose the demand $D$ is a random variable, and $d$ is its realization; so, is natural think about optimize the expected cost $\mathbb{E}[F(x, D)]$, and consequently the corresponding optimization problem would be

$$\operatorname*{Min}_{x\geq 0} \mathbb{E}[F(x, D)] \tag{2-2}$$

Let $H(d)$ be the cumulative distribution function of $D$, and note that $H(d) = 0$ for $d < 0$. Put $f(x) = \mathbb{E}[F(x, D)] = cx + \mathbb{E}[b[D - x]_+ + h[x - D]_+]$ , so

$$\begin{aligned}
f'(x) &= c + \mathbb{E}\left\{\frac{\partial(b[D-x]_+ + h[x-D]_+)}{\partial x}\right\} \\
&= c + \int_0^x \frac{\partial(b\overbrace{[d-x]_+}^{=0} + h[x-d]_+)}{\partial x} dH(d) \\
&+ \int_x^{+\infty} \frac{\partial(b[d-x]_+ + \overbrace{h[x-d]_+)}^{=0}}{\partial x} dH(d) \\
&= c + \int_0^x \frac{\partial(h(x-d))}{\partial x} dH(d) + \int_x^{+\infty} \frac{\partial(b(d-x))}{\partial x} dH(d) \\
&= c + h\int_0^x dH(d) - b\int_x^{+\infty} dH(d) \\
&= c + h[H(x) - \overbrace{H(0)}^{=0}] - b[\overbrace{H(+\infty)}^{=1} - H(x)]
\end{aligned}$$

thus

$$f'(x) = c - b + (b + h)H(x) \tag{2-3}$$

Making $f'(x) = 0$, we have that the optimal solution of the problem (2-2) is

$$\bar{x} = H^{-1}\left(\frac{b-c}{b+h}\right) \tag{2-4}$$

The example above belong to a general class of decision-making problems under uncertainty, known as two-stage stochastic linear problems with recourse. In that class, the decision process is made in two stage. In the first one, all decisions that must be taken before the realization of uncertain data, are taken. In the second one, decisions (recourse actions) that can wait to uncertain data becomes known, are considered. Examples of first-stage decisions are: (i) the wealth to be invested in a set of stocks market; (ii) the raw material quantity to make some product; (iii) the expansion option for some company. According to that, examples of second-stage decisions would be: (i) after the investment, we can wait to watch the market behaviour and then decide if buy more, sell or keep the same portfolio; (ii) after buy the raw material we can wait to know the demand and decide if buy more or maybe sell the excedent amount; (iii) if the company decided expand, it is possible to wait to know the true expansion cost and after that decide if this will be charged in the product price or not.

## 2.2 Two-Stage Stochastic Linear Model

Some basic properties of the two-stage model are presented.

### 2.2.1 Basic Properties

In general, the two-stage stochastic linear programming problem is write as

$$\begin{aligned} \operatorname*{Min}_{x\in\mathbb{R}^n} \quad & c^t x + \mathbb{E}[Q(x,\xi)] \\ \text{s.t.} \quad & Ax = b, \quad x \geq 0, \end{aligned} \tag{2-5}$$

where $Q(x,\xi)$ is the optimal value of the second stage problem

$$\begin{aligned} \operatorname*{Min}_{y\in\mathbb{R}^m} \quad & q^t y \\ \text{s.t.} \quad & Tx + Wy = h, \quad y \geq 0. \end{aligned} \tag{2-6}$$

Here $\xi := (q,h,T,W)$ is the data of the second stage problem. It is worth clarifying that part or maybe whole vector $\xi$ can be random and the expectation value is taken with respect to its probability distribution [1] (Shapiro et al., 2009).

If $\xi := (h,T)$ represents the unknown data for the problem, then $Q(x,\xi)$ it is convex for all $\xi$. To see this, let us consider the dual problem of (2-6) given by

$$\begin{aligned} \operatorname*{Min}_{\pi} \quad & \pi^t(h - Tx) \\ \text{s.t.} \quad & W^t\pi \leq q. \end{aligned} \tag{2-7}$$

By strong duality, if the problems (2-6) and (2-7) are both feasible and bounded, theirs optimal values are the same.

Define the function

$$s_q(\chi) := \inf\{q^t y | Wy = \chi, y \geq 0\}, \tag{2-8}$$

[1]The notation $\mathbb{E}_\xi[\cdot]$ will be used to enfasize that the expected value is taken respect to the $\xi$ distribution function.

with $W$ and $q$ as known parameters. Note that $Q(x,\xi) = s_q(h - Tx)$. By strong duality, if the set

$$\Pi(q) := \{\pi | W^t \pi \leq q\} \tag{2-9}$$

is nonempty, then

$$s_q(\chi) = \sup_{\pi \in \Pi(q)} \pi^t \chi \tag{2-10}$$

by other hand, $\Pi(q)$ thus define, it is a convex and closed set. In that sense, we have the follow proposition

**Proposition 2.2.1.** *The function $Q(\cdot, \xi)$ is convex for any $\xi := (h, T)$.*

***Proof*.** *Let be $\alpha \in [0,1]$, so*

$$\begin{aligned} s_q(\alpha\chi_1 + (1-\alpha)\chi_2) &= \mathop{sup}_{\pi \in \Pi(q)} \pi^t(\alpha\chi_1 + (1-\alpha)\chi_2) \\ &= \mathop{sup}_{\pi \in \Pi(q)} \{\alpha\pi^t\chi_1 + (1-\alpha)\pi^t\chi_2\} \\ &\leq \alpha \cdot \mathop{sup}_{\pi \in \Pi(q)} \pi^t\chi_1 + (1-\alpha) \cdot \mathop{sup}_{\pi \in \Pi(q)} \pi^t\chi_2, \end{aligned}$$

*thus*

$$s_q(\alpha\chi_1 + (1-\alpha)\chi_2) \leq \alpha s_q(\chi_1) + (1-\alpha)s_q(\chi_2), \quad \forall \alpha \in [0,1]. \tag{2-11}$$

*Hence $s_q(\chi)$ is a convex function for all $\chi$. According to this, since $Q(x,\xi) = s_q(h - Tx)$, $Q(\cdot,\xi)$ is a convex function also.* □

## 2.3 Formulation Problem to Discrete Distributions (a Motivator Case)

Suppose the support $\Xi \subseteq \mathbb{R}^d$ of $\xi$ is finite. That means $\xi$ has a finite number of realizations $\xi_n = (q_n, h_n, T_n, W_n)$ with probabilities $p_n > 0$, respectively, for all $n = 1, \ldots, N$. Then

$$\mathbb{E}[Q(x,\xi)] = \sum_{n=1}^{N} p_n Q(x, \xi_n). \tag{2-12}$$

For a given $x$, the expectation $\mathbb{E}[Q(x,\xi)]$ is equal to the optimal value of the linear programming problem

$$\begin{aligned} \mathop{\text{Min}}_{y_1,\cdots,y_N} \quad & \sum_{n=1}^{N} p_n q_n^t y_n \\ \text{s.t.} \quad & T_n x + W_n y_n = h_n, \quad n = 1, \ldots, N \\ & y_n \geq 0, \quad n = 1, \ldots, N. \end{aligned} \tag{2-13}$$

So, in the two-stage problem (2-5), it is possible to mix the first stage problem with the second one

$$\begin{aligned} \underset{x,y_1,\dots,y_N}{\text{Min}} \quad & c^t x + \sum_{n=1}^{N} p_n q_n^t y_n \\ \text{s.t.} \quad & T_n x + W_n y_n = h_n, \quad n = 1, \dots, N, \\ & Ax = b, \\ & x \geq 0,\ y_n \geq 0, \quad n = 1, \dots, N. \end{aligned} \tag{2-14}$$

## 2.4 Scenario Model and Sample Average Approximation (SAA)

In the probability theory framework, the random variable $\xi$ maps points of some sample space $\Omega$ into a subset of $\mathbb{R}^d$. We call the scenario of each realization of $\xi$, i.e. its value for each sample point. For problems in the real world, the sample space $\Omega$ could be very large and consequetly, the realizations set of $\xi$ result is infinite.

Inspired by the two-stage formulation problem for discrete distributions, the idea of *Sample Average Approximation*[2] is consider an smaller space $\Omega_N = \{\omega_1, \dots, \omega_N\}$, and create a finite set of scenarios $\{\xi_1 = \xi(\omega_1), \dots, \xi_N = \xi(\omega_N)\}$ uniformly distributed according to the distribution of $\xi$ via Monte Carlo simulation [3]

$$\begin{aligned} \underset{x,y_1,\dots,y_N}{\text{Min}} \quad & c^t x + \frac{1}{N}\sum_{n=1}^{N} q(\omega_n)^t y_n \\ \text{s.t.} \quad & T(\omega_n) x + W(\omega_n) y_n = h(\omega_n), \quad \forall n = 1, \dots, N, \\ & Ax = b, \\ & x \geq 0,\ y_n \geq 0, \quad \forall n = 1, \dots, N. \end{aligned} \tag{2-15}$$

The goal is that the optimal value and optimal solution of (2-15), converge to their counterparts for the true problem (2-5), increasing the scenarios number.

The SAA method is a natural way to solve the original problem and it has been broadly studied in the literature; (Kleywegt et al., 2002) is classic

[2]The following applies to any stochastic programming problem

$$\underset{x \in X}{\text{Min}} \{f(x) := \mathbb{E}[F(x,\xi)]\}.$$

[3]Those realizations can be viewed also as historical data of $N$ observations of $\xi$.

work where we can find a detailed explanation about its implementation and statistical properties.

## 2.5 Statistical Properties of SAA Estimators

What are presented in this section is strongly based on the chapter 5 (**p. 155 - 163**), with the same name, of (Shapiro et al., 2009).

Let us consider the general stochastic programming problem

$$\operatorname*{Min}_{x\in X} \{f(x) := \mathbb{E}[F(x,\xi)]\}, \tag{2-16}$$

where $X \subset \mathbb{R}^n$, is closed, $\xi$ is a random variable which probability distribution is supported on a set $\Xi \subset \mathbb{R}^d$, and $F : X \times \Xi \longrightarrow \mathbb{R}$ is a function that can be viewed as a recourse action in the two-stage model.

Let $\xi_1, \ldots, \xi_N$, $N$ be realizations of $\xi$. So, according to the above, the SAA formulation to (2-16) is given by

$$\operatorname*{Min}_{x\in X} \left\{ \hat{f}_N(x) := \frac{1}{N}\sum_{n=1}^{N}[F(x,\xi_n)] \right\}. \tag{2-17}$$

It is important to say that SAA estimator depends on the sample and consequently it is random; each realization of the sample is taken with $1/N$ of probability and also all of them have the same marginal distribution (the $\xi$ distribution). Thus, $\{\xi_n\}_{n=1}^N$ represent a family of $N$ random variables independent and identically distributed ($i.i.d$).

In that sense, $\hat{f}_N(x)$ is an ununbiased estimator of $f(x)$, since

$$\begin{aligned}\mathbb{E}[\hat{f}_N(x)] &= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[F(x,\xi_n)] = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[F(x,\xi)]\\ &= \mathbb{E}[F(x,\xi)] = f(x)\end{aligned}$$

By other hand, by the Law of Large Numbers we have that, under certain regularity conditions, $\hat{f}_N(x)$ converges pointwise with probability 1 (w.p.1), to $f(x)$ as $N \longrightarrow \infty$, since for any $x \in X$, if $\{\xi_n = \xi_n(\omega)\}_{n\in\mathbb{N}}$ is a sequence of random variables ($i.i.d$) defined in some probability space $\Omega$ with probability measure $P$, then $\{F(x,\xi_n)\}_{n\in\mathbb{N}}$ is also ($i.i.d$) and

$$P\left(\underbrace{\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}F(x,\xi_n)}_{\hat{f}_N(x)}=\overbrace{\mathbb{E}[F(x,\xi)]}^{f(x)}\right)=1, \qquad (2\text{-}18)$$

it holds. (See. A.1).

### 2.5.1 Consistency of SAA Estimators

It is said that an estimator $\hat{\theta}_N$ of a parameter $\theta$ is *consistent* if $\hat{\theta}_N \longrightarrow \theta$ as $N \longrightarrow \infty$ w.p.1. The SAA estimator is consistent.

Let $\hat{\vartheta}_N$ be the optimal value of the SAA estimator (2-17), and by $\vartheta^*$ the optimal value of the true problem (2-16), respectively. It is clear that $\hat{\vartheta}_N \leq \hat{f}_N(x)$, but also, if (2-18) it holds, it is true that $\limsup_{N\longrightarrow\infty}\hat{\vartheta}_N \leq f(x)$ w.p.1, since for any $k$ and $N$ we have

$$\sup_{k\geq N}\hat{\vartheta}_k \leq \hat{f}_N(x)$$

and

$$P\left(\overbrace{\lim_{N\to\infty}\sup_{k\geq N}\hat{\vartheta}_k}^{\mathit{lim\,sup}_{N\longrightarrow\infty}\hat{\vartheta}_N} \leq \underbrace{\lim_{N\to\infty}\hat{f}_N(x)}_{f(x)}\right) = P\left(\limsup_{N\longrightarrow\infty}\hat{\vartheta}_N \leq f(x)\right)$$

$$= P\left(\limsup_{N\longrightarrow\infty}\hat{\vartheta}_N \leq \underbrace{\inf_{x\in X} f(x)}_{\vartheta^*}\right)$$

$$= P\left(\limsup_{N\longrightarrow\infty}\hat{\vartheta}_N \leq \vartheta^*\right) = 1.$$

Additionally, we have the following proposition:

**Proposition 2.5.1.** *Suppose that $\hat{f}_N(x)$ converges to $f(x)$ w.p.1, as $N \longrightarrow \infty$, uniformly on $X$. Then $\hat{\vartheta}_N$ converges to $\vartheta^*$ w.p.1 as $N \longrightarrow \infty$.*

***Proof.*** *The uniform convergence w.p.1 of $\hat{f}_N(x) = \hat{f}_N(x,\omega)$ to $f(x)$ means that for $\epsilon > 0$ and $\omega \in \Omega$ there is $N^* = N^*(\omega)$ so that*

$$\mathit{sup}_{x\in X}|\hat{f}_N(x,\omega) - f(x)| \leq \epsilon$$

*Since*

$$|\hat{f}_N(x,\omega) - f(x)| \leq \sup_{x\in X} |\hat{f}_N(x,\omega) - f(x)|$$

*we have*

$$\hat{f}_N(x) - f(x) \leq \epsilon \quad and \quad f(x) - \hat{f}_N(x) \leq \epsilon$$

*in both cases*

$$\underbrace{\inf_{x\in X} \hat{f}_N(x,\omega)}_{\hat{\vartheta}_N} - \underbrace{\inf_{x\in X} f(x)}_{\vartheta^*} \leq \epsilon \quad and \quad \underbrace{\inf_{x\in X} f(x)}_{\vartheta^*} - \underbrace{\inf_{x\in X} \hat{f}_N(x,\omega)}_{\hat{\vartheta}_N} \leq \epsilon$$

*It follows then that* $|\hat{\vartheta}_N(\omega) - \vartheta^*| \leq \epsilon$ *for all* $N \geq N^*$. □

### 2.5.2
### Asymptotics of the SAA Optimal Value

As mentioned earlier, a SAA estimator is random and for that reason is important measure the error magnitude (dispersion) for a given sample.

If we suposse that sample is $(i.i.d)$, then for a fix point $x \in X$, we have that estimator $\hat{f}_N(x)$ of $f(x)$ is unbiased and has variance $\sigma^2(x)/N$, where $\sigma^2(x) := \mathbb{V}\mathrm{ar}[F(x,\xi)]$ is supposed finite. Moreover, by the Central Limit Theorem (See. A.2) we have that

$$N^{1/2}[\hat{f}_N(x) - f(x)] \Longrightarrow Y_x, \tag{2-19}$$

where "$\Longrightarrow$" denote convergence in distribution and $Y_x$ has a normal distribution with mean 0 and variance $\sigma^2(x)$, detonated by $Y_x \sim \mathcal{N}(0, \sigma^2(x))$. Thats means $\hat{f}_N(x)$ has *asymptotically normal* distribution.

This leads to the following $100(1-\alpha)\%$ confidence interval for $f(x)$:

$$\left[\hat{f}_N(x) - \frac{z_{\alpha/2}\hat{\sigma}(x)}{\sqrt{N}}, \hat{f}_N(x) + \frac{z_{\alpha/2}\hat{\sigma}(x)}{\sqrt{N}}\right], \tag{2-20}$$

where $z_{\alpha/2} := \Phi^{-1}(1-\alpha/2)$, and [4]

$$\hat{\sigma}^2(x) := \frac{1}{N-1}\sum_{n=1}^{N}[F(x,\xi_n) - \hat{f}_N(x)]^2 \tag{2-21}$$

is the sample variance estimate of $\sigma^2(x)$. That is, the error of estimation of $f(x)$ is (stochastically) of order $O_p(N^{-1/2})$.

Let us consider the optimal value $\hat{\vartheta}_N$ of SAA problem (2-17). It is clear that for any $x' \in X$ the inequality $\hat{f}_N(x') \geq \inf_{x\in X} \hat{f}_N(x)$ holds. By taking the

[4]Here $\Phi(\cdot)$ denotes the cdf of standard normal distribution

expected value of both sides of this inequality and minimizing the left hand side over all $x' \in X$ we obtain

$$\inf_{x \in X} \mathbb{E}[\hat{f}_N(x)] \geq \mathbb{E}\left[\inf_{x \in X} \hat{f}_N(x)\right]. \tag{2-22}$$

Since $\mathbb{E}[\hat{f}_N(x)] = f(x)$, it follows that $\vartheta^* \geq \mathbb{E}[\hat{\vartheta}_N]$. In fact, typically, $\mathbb{E}[\hat{\vartheta}_N]$ is strictly less than $\vartheta^*$, i.e., $\hat{\vartheta}_N$ is a downwards biased estimator of $\vartheta^*$. As the following result shows, this bias decrease monotonically with increase of the sample size $N$.

**Proposition 2.5.2.** *Let $\hat{\vartheta}_N$ be the optimal value of SAA problem (2-17), and suppose that the sample is (i.i.d). Then $\mathbb{E}[\hat{\vartheta}_N] \leq \mathbb{E}[\hat{\vartheta}_{N+1}] \leq \hat{\vartheta}^*$ for any $N \in \mathbb{N}$.*

***Proof*.** *We seen above that $\mathbb{E}[\hat{\vartheta}_N] \leq \vartheta^*$ for any $N \in \mathbb{N}$. Put*

$$\begin{aligned}
\hat{f}_{N+1}(x) &= \frac{1}{N+1}\sum_{n=1}^{N+1} F(x,\xi_n) \\
&= \frac{1}{N+1}\left(\frac{1}{N}(N \cdot F(x,\xi_1) + N \cdot F(x,\xi_2) + \cdots + N \cdot F(x,\xi_{N+1}))\right) \\
&= \frac{1}{N+1} \times \frac{1}{N} \times (\underbrace{F(x,\xi_2) + F(x,\xi_3) + \cdots + F(x,\xi_{N+1})}_{\sum_{m\neq 1} F(x,\xi_m)} \\
&\quad + \underbrace{F(x,\xi_1) + F(x,\xi_3) + \cdots + F(x,\xi_{N+1})}_{\sum_{m\neq 2} F(x,\xi_m)} + \cdots \\
&\quad \cdots + \underbrace{F(x,\xi_1) + F(x,\xi_2) + \cdots + F(x,\xi_N)}_{\sum_{m\neq N+1} F(x,\xi_m)}) \\
&= \frac{1}{N+1}\sum_{n=1}^{N+1}\left[\frac{1}{N}\sum_{m\neq n} F(x,\xi_m)\right].
\end{aligned}$$

*Since the sample is (i.i.d) we have that*

$$\mathbb{E}\left[\underset{x \in X}{inf} \frac{1}{N}\sum_{m\neq n} F(x,\xi_m)\right] = \mathbb{E}[\hat{\vartheta}_N]. \tag{2-23}$$

*It follows that*

$$\begin{aligned}
\mathbb{E}[\hat{\vartheta}_{N+1}] &= \mathbb{E}\left[\inf_{x\in X} \hat{f}_{N+1}(x)\right] \\
&= \mathbb{E}\left[\inf_{x\in X} \frac{1}{N+1}\sum_{n=1}^{N+1}\left(\frac{1}{N}\sum_{m\neq n} F(x,\xi_m)\right)\right] \\
&\geq \mathbb{E}\left[\frac{1}{N+1}\sum_{n=1}^{N+1}\left(\inf_{x\in X}\frac{1}{N}\sum_{m\neq n} F(x,\xi_m)\right)\right] \\
&= \frac{1}{N+1}\sum_{n=1}^{N+1}\mathbb{E}\left[\inf_{x\in X}\frac{1}{N}\sum_{m\neq n} F(x,\xi_m)\right] \\
&= \frac{1}{N+1}\sum_{n=1}^{N+1}\mathbb{E}[\hat{\vartheta}_N] = \mathbb{E}[\hat{\vartheta}_N],
\end{aligned}$$

*which completes the proof.* □

## 2.6 The Farmer's Problem

One farmer specializes in growing wheat, corn and beet. He has $500km^2$ of land and he must decide how much land would be allocated to grow each crop. By other hand, he must attend several restrictions related with his plantation. First, he must have at least 200 tons ($T$) of wheat, and $240T$ of corn, to feed his livestock. Those quantities can be obtained by own plantation or buying them in the market. The buying price per wheat ton is $\$238/T$, and per corn ton is $\$210/T$. Additionaly, every excess can be sold in the market at price of $\$170/T$ for wheat and $\$150/T$ for corn. Another important restriction is related with the beet sale. By legislative imposition the sale price per beet ton is fixed in $\$36/T$ for the first $6000T$ sold. After that, the sale price becomes $\$10/T$. We assume that plantation cost to each crop is: $\$150/km^2$, for wheat; $\$230/km^2$ for corn and $\$260/km^2$ for beet.

There is uncertainty in the land productivity. The farmer does not know the efficiency of his land for each crop.

In that context we will present the two-stage formulation of the Farmer's problem. Note that there exist three uncertainty fonts (land productivity) and three first-stage decisions (land amount allocated). Moreover, we have the amounts sold and the amounts bought as second-stage decisions:

| First-Stage Variables | |
|---|---|
| $x_T$ : | Land amount allocated for wheat (in $km^2$) |
| $x_M$ : | Land amount allocated for corn (in $km^2$) |
| $x_B$ : | Land amount allocated for beet (in $km^2$) |

| Uncertainty | |
|---|---|
| $\xi_T$ : | Random variable that represents the land productivity for wheat (in $T/km^2$) |
| $\xi_M$ : | Random variable that represents the land productivity for corn (in $T/km^2$) |
| $\xi_B$ : | Random variable that represents the land productivity for beet (in $T/km^2$) |

| Second-Stage Variables | |
|---|---|
| $y_T$ : | Wheat amount bought (in $T$) |
| $y_M$ : | Corn amount bought (in $T$) |
| $w_T$ : | Wheat amount sold (in $T$) |
| $w_M$ : | Corn amount sold (in $T$) |
| $w_{B_1}$ : | Beet amount sold at favorable price (in $T$) |
| $w_{B_2}$ : | Beet amount sold at unfavorable price (in $T$) |

Since the farmer wants to minimize his plantation cost, the two-stage stochastic linear programming problem is:

$$
\begin{aligned}
\min_{x} \quad & 150x_T + 230x_M + 260x_B + \mathbb{E}[Q(x,\xi)] \\
\text{s.t.} \quad & x_T + x_M + x_B \leq 500, \\
& x_T, x_M, x_B \geq 0,
\end{aligned}
\tag{2-24}
$$

where

$$
\begin{aligned}
Q(x,\xi) = \min_{y,w} \quad & 238y_T + 210y_M - 170w_T - 150w_M - 36w_{B_1} - 10w_{B_2} \\
\text{s.t.} \quad & \xi_T x_T + y_T - w_T \geq 200, \\
& \xi_M x_M + y_M - w_M \geq 240, \\
& \xi_B x_B - w_{B_1} - w_{B_2} \geq 0, \\
& y_T, y_M \geq 0, \quad w_T, w_M, w_{B_1}, w_{B_2} \geq 0.
\end{aligned}
\tag{2-25}
$$

To simplify the problem, we assume that productivity land for each crop, is represented by independent random variables with uniform distribution, which cumulative distribution functions are equal to

$$F_{\xi_T}(t) = \begin{cases} t-2, & 2 \leq t \leq 3 \\ 1, & t > 3 \\ 0, & t < 2 \end{cases} \qquad F_{\xi_M}(t) = \begin{cases} \frac{t-2.4}{1.2}, & 2.4 \leq t \leq 3.6 \\ 1, & t > 3.6 \\ 0, & t < 2.4 \end{cases}$$

$$F_{\xi_B}(t) = \begin{cases} \frac{t-16}{8}, & 16 \leq t \leq 24 \\ 1, & t > 24 \\ 0, & t < 16 \end{cases}$$

respectively.

Solving (2-24) (which implies calculate triple integrals and solve the optimization problem), we have that the first-stage optimal solution is: $x_T^* = 135.83\,km^2$, $x_M^* = 85.07\,km^2$ and $x_T^* = 279.10\,km^2$, and the farmer profit is: \$111237. See (Birge and Louveaux, 2011).

Now we will formulate the SAA estimator for (2-24) to analyze later its convergence and compare its error magnitude, according to the explained so far.

The SAA estimator for the Farmer's problem is equal to the optimal solution of the follow linear programming problem:

$$\begin{aligned}
\underset{x,y,w}{\text{Min}} \quad & \begin{cases} 150x_T + 230x_M + 260x_B + \\ \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} \begin{cases} 238y_T(\omega_n) + 210y_M(\omega_n) - 170w_T(\omega_n) \\ -\,150w_M(\omega_n) - 36w_{B_1}(\omega_n) - 10w_{B_2}(\omega_n) \end{cases} \end{cases} \\
\text{s.t.} \quad & x_T + x_M + x_B \leq 500, \\
& x_T, x_M, x_B \geq 0, \\
& \xi_T(\omega_n)x_T + y_T(\omega_n) - w_T(\omega_n) \geq 200, \\
& \xi_M(\omega_n)x_M + y_M(\omega_n) - w_M(\omega_n) \geq 240, \\
& \xi_B(\omega_n)x_B - w_{B_1}(\omega_n) - w_{B_2}(\omega_n) \geq 0, \\
& w_{B_1}(\omega_n) \leq 6000, \\
& y_T(\omega_n), y_M(\omega_n) \geq 0, \\
& w_T(\omega_n), w_M(\omega_n), w_{B_1}(\omega_n), w_{B_2}(\omega_n) \geq 0 \quad \forall n = 1, 2, \ldots, N.
\end{aligned} \tag{2-26}$$

It is worth say it that the large number of scenarios implies large scale.

# 3
# The Partition-based Method

## 3.1
## Notation and Preliminary Definitions

In this work, we assume the following notation and definitions:

We start with a probability space $(\Omega, \mathcal{B}, P)$, where $\mathcal{B}$ is a Borel sigma-field on $\Omega$, and $P$ is a probability measure. Also we consider the random vector $\xi : \Omega \longrightarrow \mathbb{R}^d$.

By computational reasons, we will assume an alternative definition of partition.

**Definition 3.1.1.** *The set $\mathcal{I} = \{I_k\}_{k=1}^n$ of cells $I_k$, is a partition of the support $\Xi \subset \mathbb{R}^d$ of $\xi$ if:*

1. $P\left(\bigcap_{k \in K} I_k\right) = 0, \quad \forall K \subseteq \{1, \ldots, n\},$
2. $\bigcup_{k=1}^{n} I_k = \Xi.$

**Definition 3.1.2.** *We say that partition $\mathcal{J} = \{J_{k'}\}_{k'=1}^m$ refines a partition $\mathcal{I} = \{I_k\}_{k=1}^n$, if $m > n$ and if $\forall k' = 1, \ldots, m \; \exists k$ such that $J_{k'} \subseteq I_k$.*

**Definition 3.1.3.** *$\xi^{(k)} \in \mathbb{R}^d$ is an extreme point of $I_k$, if $\xi^{(k)} = \alpha \cdot \xi_1 + (1-\alpha) \cdot \xi_2$ holds for some $\xi_1, \xi_2 \in I_k$ such that $\xi_1 \neq \xi_2$ and $\alpha \in [0,1]$, then $\alpha = 0$ or $\alpha = 1$.*

The support of the random vector $\xi = (\xi_i)_{i=1}^d$ is of the form:

$$\Xi = \prod_{i=1}^{d} \Xi_i, \tag{3-1}$$

$\Xi_i$ is the support of the random variable $\xi_i : \Omega \longrightarrow \mathbb{R}$, for all $i = 1, \ldots, d$.

In the Farmer's problem case, we have just three dimensions (wheat (T), corn (M) and beet (B)), so

$$I_k = [\xi_T^{(k,-)}, \xi_T^{(k,+)}] \times [\xi_M^{(k,-)}, \xi_M^{(k,+)}] \times [\xi_B^{(k,-)}, \xi_B^{(k,+)}], \tag{3-2}$$

Additionally, we denote the mean $\bar{\xi}^{(k)} = \mathbb{E}[\xi | \xi \in I_k]$ of $\xi$ in the cell $k$ by:

$$\bar{\xi}^{(k)} = (\bar{\xi}_T^{(k)}, \bar{\xi}_M^{(k)}, \bar{\xi}_B^{(k)})^t,$$

In the case of a uniform distribution we would have

$$\bar{\xi}_i^{(k)} = \frac{\xi_i^{(k,-)} + \xi_i^{(k,+)}}{2}, \quad \forall i \in \{T, M, B\} \tag{3-3}$$

According to the above, the set of extreme points of $I_k$, which we denote by $E(I_k)$, could be write explicitly as:

$$E(I_k) = \{(\xi_T^{(k,a)}, \xi_M^{(k,b)}, \xi_B^{(k,c)}) | a, b, c \in \{-, +\}\} \tag{3-4}$$

In the case $n = 1$, we have

$$I_1 = \Xi = [\xi_T^-, \xi_T^+] \times [\xi_M^-, \xi_M^+] \times [\xi_B^-, \xi_B^+].$$

For a uniform distribution

$$\bar{\xi}^{(1)} = \bar{\xi} = \left( \bar{\xi}_T = \frac{\xi_T^- + \xi_T^+}{2}, \bar{\xi}_M = \frac{\xi_M^- + \xi_M^+}{2}, \bar{\xi}_B = \frac{\xi_B^- + \xi_B^+}{2} \right)^t.$$

$\chi_{I_k}(\xi)$ will denote the indicator function of the set $I_k$ defined by

$$\chi_{I_k}(\xi) = \begin{cases} 1 & if \quad \xi \in I_k \\ 0 & if \quad \xi \notin I_k \end{cases}$$

## 3.2 Motivation

Give a partition $\{I_k\}_{k=1}^n$ of $\Xi$, by the total probability law (see A.5), the expected value $\mathbb{E}[Q(x, \xi)]$ can be expressed as:

$$\begin{aligned} \mathbb{E}[Q(x,\xi)] &= \sum_{k=1}^{n} \mathbb{E}[Q(x,\xi) | \xi \in I_k] \cdot P(\xi \in I_k) \\ &= \sum_{k=1}^{n} \mathbb{E}[Q(x,\xi)\chi_{I_k}(\xi)], \end{aligned} \tag{3-5}$$

By other hand, we have the following results:

$$\begin{aligned} \mathbb{E}[Q(x,\xi)\chi_{I_k}(\xi)] &= \int_{\Omega} Q(x, \xi(\omega))\chi_{I_k}(\xi(\omega))dP(\omega) \\ &= \int_{I_k} Q(x, \xi(\omega))dP(\omega) \\ &\leq \int_{I_k} \underset{\xi \in I_k}{Max}\, Q(x,\xi)dP(\omega) \\ &= \underset{\xi \in I_k}{Max}\, Q(x,\xi) \int_{I_k} dP(\omega) \\ &= \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in I_k), \end{aligned}$$

then

$$\mathbb{E}[Q(x,\xi)\chi_{I_k}(\xi)] \leq \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in I_k),$$

In this way it holds that

$$\mathbb{E}[Q(x,\xi)] = \sum_{k=1}^{n} \mathbb{E}[Q(x,\xi)\chi_{I_k}(\xi)] \leq \sum_{k=1}^{n} \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in I_k). \tag{3-6}$$

Additionally, if $Q(x,\xi)$ is a convex function in $\xi$ for any $x$, by Jensen's inequality (see A.4), it holds that

$$Q(x, \mathbb{E}[\xi | \xi \in I_k]) \leq \mathbb{E}[Q(x,\xi) | \xi \in I_k],$$

so

$$\sum_{k=1}^{n} Q(x, \mathbb{E}[\xi | \xi \in I_k]) \cdot P(\xi \in I_k) \leq \mathbb{E}[Q(x,\xi)]. \tag{3-7}$$

The equations (3-6) and (3-7) show that partitioning the support $\Xi$ considering the first conditional moment, we can find deterministic bounds for the true (continously distributed) problem (5-1).

So, the remainder problem is finding an optimal way to split the support $\Xi$, to reduce the bounds difference gap and converge to the optimal solution of the true problem.

The partitioning problem consist of finding a finite set of probabilities that fits the underlying distribution of $\xi$ and allow discrete the expected operator to find an approximate solution in the sense of numerical integration. The common procedures of the exist approximation methods are based on the Jensen inequality and Edmundson-Madansky inequality to set a lower and upper bound, respectively, for the case which the recourse function is convex in the uncertain variable (Birge and Louveaux, 2011).

The Edmundson-Madansky inequality (see A.8) illustrates that if $\xi \longmapsto Q(x,\xi)$ is convex, and we consider the set of extreme points of the convex hull of $\Xi$, denoted by $ext\Xi$ , and the sigma algebra $\mathcal{E}$ of all subsets of $ext\Xi$, then if we can express any $\xi \in \Xi$ as a convex combination (average) of extreme points, i.e. if $\phi(\xi, \cdot)$ is a probability measure on $(ext\Xi, \mathcal{E})$ such that

$$\int_{e \in ext\Xi} e \cdot \phi(\xi, de) = \xi, \quad \forall \xi \in \Xi, \tag{3-8}$$

then

$$\mathbb{E}[Q(x,\xi)] \leq \int_{e \in ext\Xi} Q(x,e)\lambda(de), \tag{3-9}$$

where $\lambda$ is the probability measure on $\mathcal{E}$ defined by

$$\lambda(A) = \int_{\Omega} \phi(\xi(\omega), A)P(d\omega) \quad \forall A \in \mathcal{E} \tag{3-10}$$

By the condition (3-8), it holds:

$$\begin{aligned}
\int_{\Omega} \xi(\omega)P(d\omega) &= \int_{\Omega} \left( \int_{e \in ext\Xi} e \cdot \phi(\xi(\omega), de) \right) P(d\omega) \\
&= \int_{e \in ext\Xi} e \int_{\Omega} \phi(\xi(\omega), de)P(d\omega) \\
&= \int_{e \in ext\Xi} e\lambda(de).
\end{aligned}$$

## 3.3 Method

The partition-based method consists in solve the sequential refinement problem, spliting the support of the distribution and finding the upper and lower solution for the given partition, until obtain a gap small enough to garant that the estimated solution is convergent.

---

**Algorithm 1** Solve the Bounding Problem

---

**Require:** $\mathcal{I} = \{I_k\}_{k=1}^n$, $\epsilon$ stopping criteria.

**Ensure:** $\vartheta_L^*$ optimal value for the lower bound problem, $\vartheta_U^*$ optimal value for the upper bound problem.

**Step 1**

Solve

$$\vartheta_U^* = \begin{cases} \underset{x}{\text{Min}} & c^t x + \sum_{k=1}^{n} \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in I_k) \\ \text{s.t.} & x \in X. \end{cases}$$

Solve

$$\vartheta_L^* = \begin{cases} \underset{x}{\text{Min}} & c^t x + \sum_{k=1}^{n} Q(x, \mathbb{E}[\xi | \xi \in I_k]) \cdot P(\xi \in I_k) \\ \text{s.t.} & x \in X. \end{cases}$$

**if** $\vartheta_U^* - \vartheta_L^* \leq \epsilon$ **then**

  **stop**

**else**

  go to **Algorithm 1** or **Algorithm 2** with $\mathcal{I} = \{I_k\}_{k=1}^n$, and return to **Step 1**

**end if**

---

### 3.3.1 Convergence of the Method

In this section we will present the proof about the convergence of the partition-based method.

***Proof.*** *(**Upper bound**) Define* $\mathscr{F} = \{\mathcal{I} = \{I_k\}_{k=1}^n \,|\, \mathcal{I} \text{ is partition of } \Xi\}$. *We have by the Total Probability Law (A.5) that:*

$$\mathbb{E}[Q(x,\xi)] = \sum_{k=1}^{n} \int_{\Omega} Q(x,\xi(\omega))\chi_{I_k}(\xi(\omega))dP(\omega) \quad \forall \mathcal{I} = \{I_k\}_{k=1}^n \in \mathscr{F}, \tag{3-11}$$

*also, we saw that*

$$\sum_{k=1}^{n} \int_{\Omega} Q(x,\xi(\omega))\chi_{I_k}(\xi(\omega))dP(\omega) \leq \sum_{k=1}^{n} \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in I_k) \quad \forall \mathcal{I} = \{I_k\}_{k=1}^n \in \mathscr{F}. \tag{3-12}$$

*Given* $\mathcal{I} = \{I_k\}_{k=1}^n \in \mathscr{F}$, *put*

$$\theta(\mathcal{I}) = \sum_{k=1}^{n} \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in I_k).$$

*Consider* $\mathcal{J} = \{J_{k'}\}_{k'=1}^m$ *and* $\mathcal{I} = \{I_k\}_{k=1}^n$ *in* $\mathscr{F}$ *such that* $\mathcal{J}$ *refines the partition* $\mathcal{I}$, *then by definition* $m > n$ *and* $\forall k' \; \exists k$ *such that*

$$J_{k'} \subseteq I_k.$$

*As for all* $\mathcal{I} = \{I_k\}_{k=1}^n \in \mathscr{F}$, $int I_k \neq \emptyset \quad \forall k$ *and* $int I_k \cap int I_{k'} = \emptyset \quad \forall k \neq k'$, *this implies that for all* $k = 1, \ldots, n$, *there is a set* $\{{k'}_1^{(k)}, \ldots, {k'}_{s_k}^{(k)}\} \subset \{1, \ldots, m\}$, *such that*

$$I_k = \bigcup_{i \in \{{k'}_1^{(k)}, \ldots, {k'}_{s_k}^{(k)}\}} J_i,$$

*and* $\{{k'}_1^{(k_1)}, \ldots, {k'}_{s_{k_1}}^{(k_1)}\} \cap \{{k'}_1^{(k_2)}, \ldots, {k'}_{s_{k_2}}^{(k_2)}\} = \emptyset$ *for all* $k_1, k_1 \in \{1, \ldots, n\}$ *with* $k_1 \neq k_2$.

*Note that*

$$\{1, \ldots, m\} = \bigcup_{k=1}^{n} \{{k'}_1^{(k)}, \ldots, {k'}_{s_k}^{(k)}\}.$$

*Thus,*

$$\underset{\xi \in J_i}{Max}\, Q(x,\xi) \leq \underset{\xi \in I_k}{Max}\, Q(x,\xi), \quad \forall i \in \{{k'}_1^{(k)}, \ldots, {k'}_{s_k}^{(k)}\}, \tag{3-13}$$

*and,*

$$\sum_{i \in \{{k'}_1^{(k)}, \ldots, {k'}_{s_k}^{(k)}\}} P(\xi \in J_i) = P(\xi \in I_k), \tag{3-14}$$

*so,*

$$\begin{aligned}
\sum_{i \in \{{k'}_1^{(k)}, \ldots, {k'}_{s_k}^{(k)}\}} \underset{\xi \in J_i}{Max}\, Q(x,\xi) \cdot P(\xi \in J_i) &\leq \sum_{i \in \{{k'}_1^{(k)}, \ldots, {k'}_{s_k}^{(k)}\}} \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in J_i) \\
&= \underset{\xi \in I_k}{Max}\, Q(x,\xi) \sum_{i \in \{{k'}_1^{(k)}, \ldots, {k'}_{s_k}^{(k)}\}} P(\xi \in J_i) \\
&= \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in I_k).
\end{aligned}$$

*In that sense*

$$\begin{aligned}
\sum_{k'=1}^{m} \underset{\xi \in J_{k'}}{Max}\, Q(x,\xi) \cdot P(\xi \in J_{k'}) &= \sum_{k=1}^{n} \sum_{i \in \{{k'}_1^{(k)}, \ldots, {k'}_{s_k}^{(k)}\}} \underset{\xi \in J_i}{Max}\, Q(x,\xi) \cdot P(\xi \in J_i) \\
&\leq \sum_{k=1}^{n} \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in I_k).
\end{aligned}$$

*Therefore,*

$$\sum_{k'=1}^{m} \underset{\xi \in J_{k'}}{Max}\, Q(x,\xi) \cdot P(\xi \in J_{k'}) \leq \sum_{k=1}^{n} \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in I_k). \tag{3-15}$$

*Let* $\{\mathcal{I}^{(n)}\}_{n\geq 1} \subseteq \mathscr{F}$ *be a indexed family of partitions such that* $\mathcal{I}^{(n)} = \{I_k\}_{k=1}^{n}$ *where* $I_k$ *is a compact and convex set* $\forall k$, *and the relation:* $\mathcal{I}^{(n+1)}$ *refines the partition* $\mathcal{I}^{(n)}$, *it holds. If we define* $\theta_n := \theta(\mathcal{I}^{(n)})$, *then according to (3-15) we have that the sequence* $\{\theta_n\}_{n\geq 1}$ *is nonincreasing.*

*Furthemore, by (3-11) and (3-12), the sequence* $\{\theta_n\}_{n\geq 1}$ *is lower bounded, therefore* $\{\theta_n\}_{n\geq 1}$ *is convergent, so*

$$\theta_n \longrightarrow \underset{\{\mathcal{I}^{(n)}\}_{n\geq 1} \subseteq \mathscr{F}}{inf} \theta_n \quad as \quad n \longrightarrow \infty.$$

*By other hand, given* $\mathcal{I}^{(n)} = \{I_k\}_{k=1}^{n} \in \{\mathcal{I}^{(n)}\}_{n\geq 1}$, *if* $Q(x,\xi)$ *is continuous, as* $I_k$ *is compact, there is* $\xi^{(k,*)}$ *such that*

$$\forall k, \quad Q(x,\xi) \leq Q(x,\xi^{(k,*)}) \quad \forall \xi \in I_k.$$

*Moreover given $\epsilon > 0$, by continuity again, $\forall k$ there is an open ball $B_{\delta_k}(\xi^{(k,*)})$ with center in $\xi^{(k,*)}$ and radio $\delta_k > 0$, such that*

$$0 \leq \underset{\xi \in I_k}{Max}\, Q(x,\xi) - Q(x,\xi) < \epsilon \quad \forall \xi \in B_{\delta_k}(\xi^{(k,*)}) \cap I_k.$$

*Therefore, given $\epsilon > 0$ it is possible to find $\mathcal{I}^{(n^*)} = \{\hat{I}_k\}_{k=1}^{n^*} \in \{\mathcal{I}^{(n)}\}_{n\geq 1}$, such that $\hat{I}_k$ is the smallest compact and convex neighborhood of $\xi^{(k,*)}$ for all $k = 1, \ldots, n^*$. So,*

$$\forall k, \quad 0 \leq \underset{\xi \in \hat{I}_k}{Max}\, Q(x,\xi) - Q(x,\xi) < \epsilon \quad \forall \xi \in \hat{I}_k. \tag{3-16}$$

*Since $\forall k$ (3-16) it holds if $\xi \in \hat{I}_k$, i.e. if $\chi_{\hat{I}_k}(\xi) = 1$, then is true the following inequality:*

$$0 \leq \underset{\xi \in \hat{I}_k}{Max}\, Q(x,\xi)\chi_{\hat{I}_k}(\xi) - Q(x,\xi)\chi_{\hat{I}_k}(\xi) < \epsilon \cdot \chi_{\hat{I}_k}(\xi), \quad \forall k. \tag{3-17}$$

*Integrating respect to the probability measure $P$, we have:*

$$\begin{aligned}\underset{\xi \in \hat{I}_k}{Max}\, Q(x,\xi) \cdot \int_\Omega \chi_{\hat{I}_k}(\xi(\omega))dP(\omega) - \int_\Omega Q(x,\xi(\omega))\chi_{\hat{I}_k}(\xi(\omega))dP(\omega) \\ < \epsilon \cdot \int_\Omega \chi_{\hat{I}_k}(\xi(\omega))dP(\omega), \quad \forall k.\end{aligned}$$

*But $\int_\Omega \chi_{\hat{I}_k}(\xi(\omega))dP(\omega) = P(\xi \in \hat{I}_k)$, so we have*

$$\begin{aligned}0 \leq \sum_{k=1}^{n^*} \underset{\xi \in \hat{I}_k}{Max}\, Q(x,\xi) \cdot P(\xi \in \hat{I}_k) - \sum_{k=1}^{n^*} \int_\Omega Q(x,\xi(\omega))\chi_{\hat{I}_k}(\xi(\omega))dP(\omega) \\ < \epsilon \cdot \underbrace{\sum_{k=1}^{n^*} P(\xi \in \hat{I}_k)}_{=1}\end{aligned} \tag{3-18}$$

*Then, given $\epsilon > 0$ there is $n^* \in \mathbb{N}$, such that*

$$0 \leq \theta_n - \mathbb{E}[Q(x,\xi)] < \epsilon, \quad \forall n \geq n^*, \tag{3-19}$$

*where $n = |\mathcal{I}^{(n)}|$ is the size of the partition $\mathcal{I}^{(n)} \in \{\mathcal{I}^{(n)}\}_{n\geq 1}$ such that $\mathcal{I}^{(n)}$ refines the partition $\mathcal{I}^{(n^*)}$.*

*According to the above, if $Q(x,\xi)$ is continuous in $\xi$, we can prove that the sequence $\theta_n \longrightarrow \mathbb{E}[Q(x,\xi)]$ as $n = |\mathcal{I}^{(n)}| \longrightarrow \infty$, with $\mathcal{I}^{(n)} \in \{\mathcal{I}^{(n)}\}_{n\geq 1}$.*

*Additionally it is clear, under that hypothesis, that if*

$$\begin{aligned} \vartheta^* := Min \quad & c^t x + \mathbb{E}[Q(x,\xi)] \\ s.t. \quad & x \in X, \end{aligned} \tag{3-20}$$

*and*

$$\begin{aligned} \vartheta_n^* := Min \quad & c^t x + \sum_{k=1}^{n} \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in I_k) \\ s.t. \quad & x \in X, \end{aligned} \tag{3-21}$$

*for some $\mathcal{I}^{(n)} \in \{\mathcal{I}^{(n)}\}_{n\geq 1}$, then $\vartheta_n^* \downarrow \vartheta^*$. Which completes the proof.* $\square$

The proof for the lower bound is quite similar to the upper one.

***Proof.*** *(**Lower Bound**) If $Q(x,\xi)$ is convex in $\xi$, we saw that*

$$\begin{aligned} &\sum_{k=1}^{n} Q(x, \mathbb{E}[\xi|\xi \in I_k]) \cdot P(\xi \in I_k) \ \leq \\ &\quad \sum_{k=1}^{n} \int_{\Omega} Q(x,\xi(\omega))\chi_{I_k}(\xi(\omega))dP(\omega) \quad \forall \mathcal{I} = \{I_k\}_{k=1}^{n} \in \mathscr{F}. \end{aligned} \tag{3-22}$$

*Consider $\mathcal{J} = \{J_{k'}\}_{k'=1}^{m}$ and $\mathcal{I} = \{I_k\}_{k=1}^{n}$ in $\mathscr{F}$, such that $\mathcal{J}$ refines $\mathcal{I}$. We know that for all $k \in \{1,\dots,n\}$ there is $\{{k'}_1^{(k)},\dots,{k'}_{s_k}^{(k)}\}$ such that*

$$I_k = \bigcup_{i \in \{{k'}_1^{(k)},\dots,{k'}_{s_k}^{(k)}\}} J_i,$$

*where $\{{k'}_1^{(k_1)},\dots,{k'}_{s_{k_1}}^{(k_1)}\} \cap \{{k'}_1^{(k_2)},\dots,{k'}_{s_{k_2}}^{(k_2)}\} = \emptyset$ for all $k_1$, $k_2 \in \{1,\dots,n\}$ with $k_1 \neq k_2$, and*

$$\{1,\dots,m\} = \bigcup_{k=1}^{n} \{{k'}_1^{(k)},\dots,{k'}_{s_k}^{(k)}\}$$

*By other hand, it is true that*

$$\forall k, \quad \mathbb{E}[\xi|\xi \in I_k] \ = \sum_{i \in \{{k'}_1^{(k)},\dots,{k'}_{s_k}^{(k)}\}} \frac{\mathbb{E}[\xi|\xi \in J_i] \cdot P(\xi \in J_i)}{P(\xi \in I_k)}. \tag{3-23}$$

*Note that*

$$\begin{aligned} \sum_{i \in \{{k'}_1^{(k)},\dots,{k'}_{s_k}^{(k)}\}} \frac{P(\xi \in J_i)}{P(\xi \in I_k)} &= \frac{1}{P(\xi \in I_k)} \sum_{i \in \{{k'}_1^{(k)},\dots,{k'}_{s_k}^{(k)}\}} P(\xi \in J_i) \\ &= \frac{1}{P(\xi \in I_k)} \cdot P(\xi \in I_k) = 1. \end{aligned}$$

*Then (3-23) is a convex combination, so*

$$Q(x,\mathbb{E}[\xi|\xi\in I_k]) \le \frac{1}{P(\xi\in I_k)} \sum_{i\in\{k'^{(k)}_1,\ldots,k'^{(k)}_{s_k}\}} Q(x,\mathbb{E}[\xi|\xi\in J_i])\cdot P(\xi\in J_i).$$

*Thus,*

$$\begin{aligned}\sum_{k=1}^{n} Q(x,\mathbb{E}[\xi|\xi\in I_k])\cdot P(\xi\in I_k) \le \\ &\sum_{k=1}^{n}\sum_{i\in\{k'^{(k)}_1,\ldots,k'^{(k)}_{s_k}\}} Q(x,\mathbb{E}[\xi|\xi\in J_i])\cdot P(\xi\in J_i)\\ &= \sum_{k'=1}^{m} Q(x,\mathbb{E}[\xi|\xi\in J_{k'}])\cdot P(\xi\in J_{k'}).\end{aligned}$$

*Therefore*

$$\sum_{k=1}^{n} Q(x,\mathbb{E}[\xi|\xi\in I_k])\cdot P(\xi\in I_k) \le \sum_{k'=1}^{m} Q(x,\mathbb{E}[\xi|\xi\in J_{k'}])\cdot P(\xi\in J_{k'}). \quad (3\text{-}24)$$

*Given* $\mathcal{I}=\{I_k\}_{k=1}^{n}\in\mathscr{F}$, *define*

$$\theta(\mathcal{I}) := \sum_{k=1}^{n} Q(x,\mathbb{E}[\xi\in I_k])\cdot P(\xi\in I_k). \quad (3\text{-}25)$$

*Let* $\{\mathcal{I}^{(n)}\}_{n\ge1}\subseteq\mathscr{F}$ *be a indexed family of partitions such that* $\mathcal{I}^{(n)}=\{I_k\}_{k=1}^{n}$ *where* $I_k$ *is a compact and convex set* $\forall k$, *and the relation:* $\mathcal{I}^{(n+1)}$ *refines the partition* $\mathcal{I}^{(n)}$, *it holds. If we define* $\theta_n := \theta(\mathcal{I}^{(n)})$, *then according to (3-24) we have that the sequence* $\{\theta_n\}_{n\ge1}$ *is nondecreasing.*

*Moreover, by (3-22) the sequence* $\{\theta_n\}_{n\ge1}$ *is upper bounded, so* $\{\theta_n\}_{n\ge1}$ *is convergent, and*

$$\theta_n \longrightarrow \sup_{\{\mathcal{I}^{(n)}\}_{n\ge1}\subseteq\mathscr{F}} \theta_n \quad as \quad n\longrightarrow\infty. \quad (3\text{-}26)$$

*As* $Q(x,\xi)$ *is convex, if* $\Xi$ *is bounded and convex, then* $Q(x,\xi)$ *is continuos (see A.3). So given* $\mathcal{I}^{(n)}=\{I_k\}_{k=1}^{n}\in\{\mathcal{I}^{(n)}\}_{n\ge1}$, *and* $\epsilon>0$, *we have that:*

$$\forall k,\quad \exists\delta_k>0 \quad s.t \quad |Q(x,\xi)-Q(x,\mathbb{E}[\xi|\xi\in I_k])|<\epsilon,\quad \forall\xi\in B_{\delta_k}(\mathbb{E}[\xi|\xi\in I_k])\cap I_k.$$

*Therefore, given* $\epsilon>0$ *it is possible to find* $\mathcal{I}^{(n^*)}=\{\hat{I}_k\}_{k=1}^{n^*}\in\{\mathcal{I}^{(n)}\}_{n\ge1}$, *such that* $\hat{I}_k$ *is the smallest compact and convex neighborhood of* $\mathbb{E}[\xi|\xi \in I_k]$ *for all* $k = 1, \ldots, n^*$. *So,*

$$\forall k, \quad |Q(x,\xi) - Q(x, \mathbb{E}[\xi|\xi \in I_k])| < \epsilon, \quad \forall \xi \in \hat{I}_k. \tag{3-27}$$

*Since for all* $k \in \{1, \ldots, n^*\}$ *(3-27) it holds if* $\xi \in \hat{I}_k$, *i.e. if* $\chi_{\hat{I}_k}(\xi) = 1$, *then it is true the following inequality:*

$$|Q(x,\xi)\chi_{\hat{I}_k}(\xi) - Q(x, \mathbb{E}[\xi|\xi \in I_k])\chi_{\hat{I}_k}(\xi)| < \epsilon, \quad \forall k,$$

*so*

$$\left| \sum_{k=1}^{n^*} \int_\Omega Q(x, \xi(\omega))\chi_{\hat{I}_k}(\xi(\omega))dP(\omega) - \sum_{k=1}^{n^*} \int_\Omega Q(x, \mathbb{E}[\xi|\xi \in \hat{I}_k])\chi_{\hat{I}_k}(\xi(\omega))dP(\omega) \right|$$
$$\leq \sum_{k=1}^{n^*} \int_\Omega |Q(x, \xi(\omega)) - Q(x, \mathbb{E}[\xi|\xi \in \hat{I}_k])|\chi_{\hat{I}_k}(\xi(\omega))dP(\omega)$$
$$< \sum_{k=1}^{n^*} \int_\Omega \epsilon \cdot \chi_{\hat{I}_k}(\xi(\omega))dP(\omega) = \epsilon \cdot \sum_{k=1}^{n^*} P(\xi \in \hat{I}_k) = \epsilon.$$

*Then*

$$\theta_n \longrightarrow \mathbb{E}[Q(x,\xi)] \quad as \quad n \longrightarrow \infty,$$

*additionally if* $\vartheta^*$ *and* $\vartheta^*_n$ *are defined as (3-20) and (3-21), respectively, then* $\vartheta^*_n \uparrow \vartheta^*$. $\square$



### 3.3.2 Convex Case

In this section we will show how to solve the problem

$$\operatorname*{Max}_{\xi \in I_k} Q(x,\xi), \qquad \text{with} \qquad I_k \in \mathcal{I} = \{I_k\}_{k=1}^n,$$

when $Q(x,\xi)$ is a convex function in $\xi$ for all $x$.

In general, if $U \subset \mathbb{R}^d$ is open and convex set, the convex function $Q(x,\cdot) : U \longrightarrow \mathbb{R}$ is continuous (see A.3).

For the case where the probability distribution function of $\xi$, has a compact support $\Xi \subset \mathbb{R}^d$, we can garantee that $Q(x,\xi)$ is continuous on the convex hull of $\Xi$. In that sense, $Q(x,\xi)$ must reach its maximum and its minimum on the convex hull of $\Xi$, since the convex hull of a compact set in

$\mathbb{R}^d$ is compact; so the set of optimal solutions for the problem $\text{Max}_{\xi\in I_k} Q(x,\xi)$, is not empty, $\forall k = 1,\dots,n$.

In addition to above, by the maximum principle of convex optimization (Rockafellar, 1970), the maximum of the convex function $Q(x,\cdot) : U \longrightarrow \mathbb{R}$ on a convex and compact set $U \subset \mathbb{R}^d$, is attained on the boundary. Nonetheless, in a more general way, the maximum of the function $Q(x,\cdot)$ relative to the set $S$ where is not constant, is attained in the extreme points of $S$ or in a convex combination of them. To see this let us consider the following:

Suposse that the maximum of $Q(x,\cdot)$ relatived to $S$ is reached in $\tilde{\xi} = \alpha\xi + (1-\alpha)\xi'$ for some $\alpha \in (0,1)$ and $\xi,\xi' \in S$ with $\xi \neq \xi'$. So, by convexity we have

$$Q(x,\tilde{\xi}) \leq \alpha Q(x,\xi) + (1-\alpha)Q(x,\xi').$$

But by hypothesis, $Q(x,\xi) \leq Q(x,\tilde{\xi})$ and $Q(x,\xi') \leq Q(x,\tilde{\xi})$. As $Q(x,\xi) \neq Q(x,\tilde{\xi})$ and $Q(x,\xi') \neq Q(x,\tilde{\xi})$ (since $\xi,\xi' \in S$), then $Q(x,\xi) < Q(x,\tilde{\xi})$ and $Q(x,\xi') < Q(x,\tilde{\xi})$. Therefore

$$Q(x,\tilde{\xi}) < \alpha Q(x,\tilde{\xi}) + (1-\alpha)Q(x,\tilde{\xi}),$$

so

$$Q(x,\tilde{\xi}) < Q(x,\tilde{\xi})$$

which is a contradiction.

Follow this idea, it is possible to prove also that if $\xi'$ is an extreme point of a free-line closed convex set $U \subseteq \mathbb{R}^d$, then

$$\xi' \in \operatorname*{Argmax}_{\xi\in U} Q(x,\xi)$$

where $Q(x,\cdot) : U \longrightarrow \mathbb{R}$ is a convex function with $x$ fixed, that must be bounded above every halfline of $U \subseteq \mathbb{R}^d$. A proof of this fact is presented in (Tuy, 1998).

In accordance with the above, $Q(x,\xi)$ reach its maximum in some point of the set $E(I_k)$ [1], i.e., $E(I_k) \subseteq \text{Argmax}_{\xi\in I_k} Q(x,\xi)$, hence the following equivalence it holds:

[1] Remembering that $E(I_k)$ denote the set of extreme points of $I_k$

$$\underset{\xi\in I_k}{\mathrm{Max}}\, Q(x,\xi) = \underset{\xi\in E(I_k)}{\mathrm{Max}}\, Q(x,\xi). \tag{3-28}$$

Explicity, $\mathrm{Max}_{\xi\in E(I_k)} Q(x,\xi)$ can be write as:

$$\underset{\xi\in E(I_k)}{\mathrm{Max}}\, Q(x,\xi) = \max\{Q(x,\xi^{(k,s)})|s \in \{-,+\}^d\}, \tag{3-29}$$

where $\xi^{(k,s)}$, is the value of $\xi$ on the vertex $s$ of the cell $k$, and $x$ is fixed.

### 3.3.3 Rectangular Partition

We consider rectangular partitions since a rectangular cell is a convex and compact set, and they allow compute easier the discrete probability distribution $\{P(\xi \in I_k)\}_{k=1}^n$ used to discretize the expectation of $Q(x,\xi)$.

A rectangular cell $I_k$ has the form:

$$I_k = \prod_{i=1}^{d}[\xi_i^{(k,-)}, \xi_i^{(k,+)}], \tag{3-30}$$

where $\xi_i^{(k,-)}$, denote the lower end of the interval on the dimension $i$ of partition $k$, and $\xi_i^{(k,+)}$ denote the upper end one (see Figure 3.1).

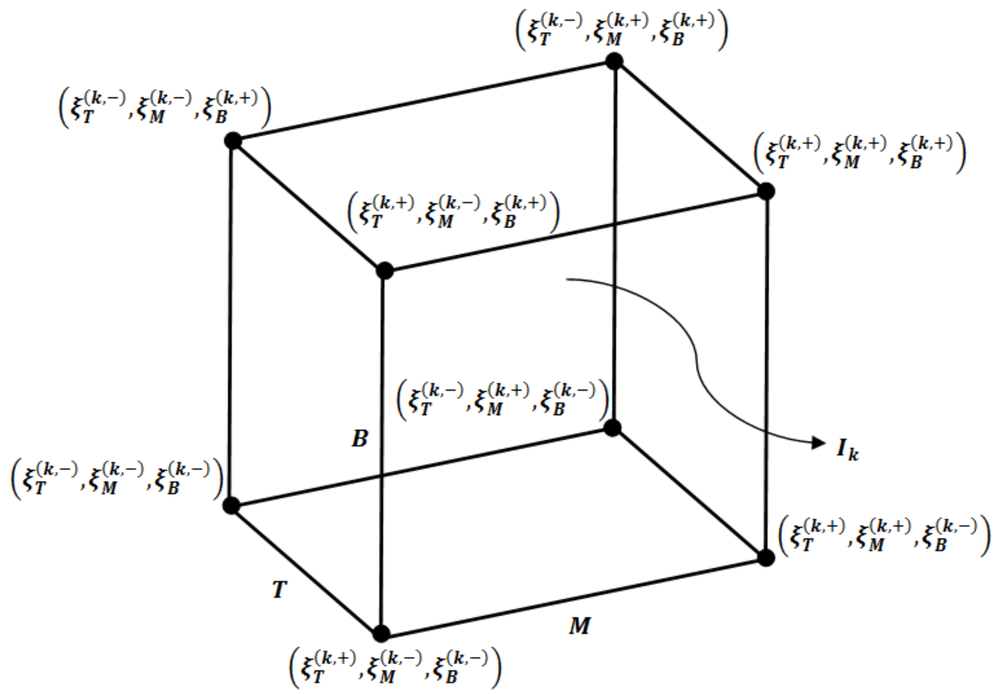



Figure 3.1: Extreme points of the cell $I_k$

# 4
# Algorithms

The three basic decisions for refining the partition from $\mathcal{I} = \{I_k\}_{k=1}^{n}$ to $\mathcal{I}' = \{I_{k'}\}_{k'=1}^{n+1}$ are to choose the cell, $I_{k*} \in \mathcal{I}$, in which to make the partition, to choose the direction in which to split $I_{k*}$, and to choose the point at which to make the split.

In this work we will assume that the cutting point is the mean of $\xi$ over the direction in which to split $I_{k*}$.

## 4.1
## Solution Algorithm with a Radom Partition Refinement (SARPR)

Here we present the first version of the partitioning algorithm. The idea is splitting uniformly the uncertainty domain (support of $\xi$ distribution function).

In the first step we just have to choose randomly the dimension $i^*$ which is going to be split to get the two initial cells. The cut is made in the mean of $\xi$ in that dimension, obtaining two subintervals; the first one is the lower end $\xi_{i^*}^{-}$ to the mean $\bar{\xi}_{i^*}$, and the second one is the mean to the upper end $\xi_{i^*}^{+}$, respectively. These subintervals are multiplied (in the sense of cartesian product) with the remainder dimensions (see Figure 4.1).

When we have more than one cell, we also have to choose randomly the cell $k^*$ that is going to be split like the original support of $\xi$.

To get the new partition that refines the previous one, we keep the others cells different from $k^*$ and add the two others that were obtained by splitting the dimension $i^*$ at the corresponding mean value of $\xi$ inside the cell $k^*$.

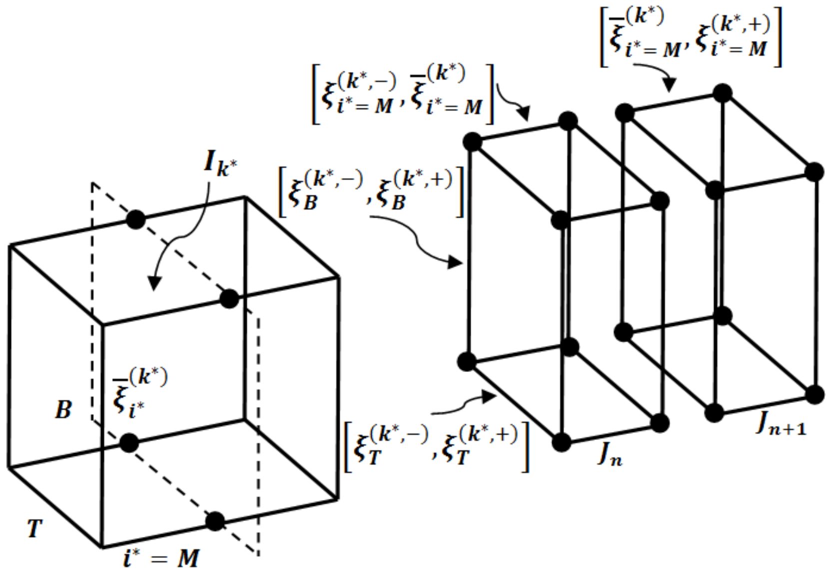



Figure 4.1: Example of refinement $i^* = M$

See also Figure 4.2.

**Algorithm 2** SARPR

**Require:** $\mathcal{I} = \{I_k\}_{k=1}^{n}$ initial partition of $\Xi$

**Ensure:** $\mathcal{J} = \{J_{k'}\}_{k'=1}^{n+1}$ such that $\mathcal{J}$ refines $\mathcal{I}$

**if** n = 1 **then**
  $r_1 \Longleftarrow$ rand
  **if** $0 \leq r_1 < 1/d$ **then**
    $i^* \Longleftarrow 1$
  **else if** $1/d \leq r_1 < 2/d$ **then**
    $i^* \Longleftarrow 2$
      $\vdots$
  **else if** $(d-2)/d \leq r_1 < (d-1)/d$ **then**
    $i^* \Longleftarrow d-1$
  **else**
    $i^* \Longleftarrow d$
  **end if**
  Put $J_1 = [\xi^-_{i^*}, \bar{\xi}_{i^*}] \times \prod_{t\in\{1,2,\dots,d\}\setminus\{i^*\}} [\xi^-_t, \xi^+_t]$ and $J_2 = [\bar{\xi}_{i^*}, \xi^+_{i^*}] \times \prod_{t\in\{1,2,\dots,d\}\setminus\{i^*\}} [\xi^-_t, \xi^+_t]$
  $\mathcal{J} = \{J_{k'}\}_{k'=1}^{2}$
**else**
  $r_1 \Longleftarrow$ rand
  $r_2 \Longleftarrow$ rand
  **for** $j = 1$ **to** $n$ **do**
    **if** $\frac{j-1}{n} \leq r_1 < \frac{j}{n}$ **then**
      $k^* \Longleftarrow j$
    **end if**
  **end for**
  **if** $0 \leq r_2 < 1/d$ **then**
    $i^* \Longleftarrow 1$
  **else if** $1/d \leq r_2 < 2/d$ **then**
    $i^* \Longleftarrow 2$
      $\vdots$
  **else if** $(d-2)/d \leq r_2 < (d-1)/d$ **then**
    $i^* \Longleftarrow d-1$
  **else**
    $i^* \Longleftarrow d$
  **end if**
  Put $k' \Longleftarrow k \; \forall k \neq k^*$, $J_{k^*} \Longleftarrow [\xi_{i^*}^{(k^*,-)}, \bar{\xi}_{i^*}^{(k^*)}] \times \prod_{t\in\{1,2,\dots,d\}\setminus\{i^*\}} [\xi_t^{(k^*,-)}, \xi_t^{(k^*,+)}]$ and
  $J_{n+1} \Longleftarrow [\bar{\xi}_{i^*}^{(k^*)}, \xi_{i^*}^{(k^*,+)}] \times \prod_{t\in\{1,2,\dots,d\}\setminus\{i^*\}} [\xi_t^{(k^*,-)}, \xi_t^{(k^*,+)}]$
  $\mathcal{J} = \{J_{k'}\}_{k'=1}^{n+1}$
**end if**

SARPR for the Farmer's Problem

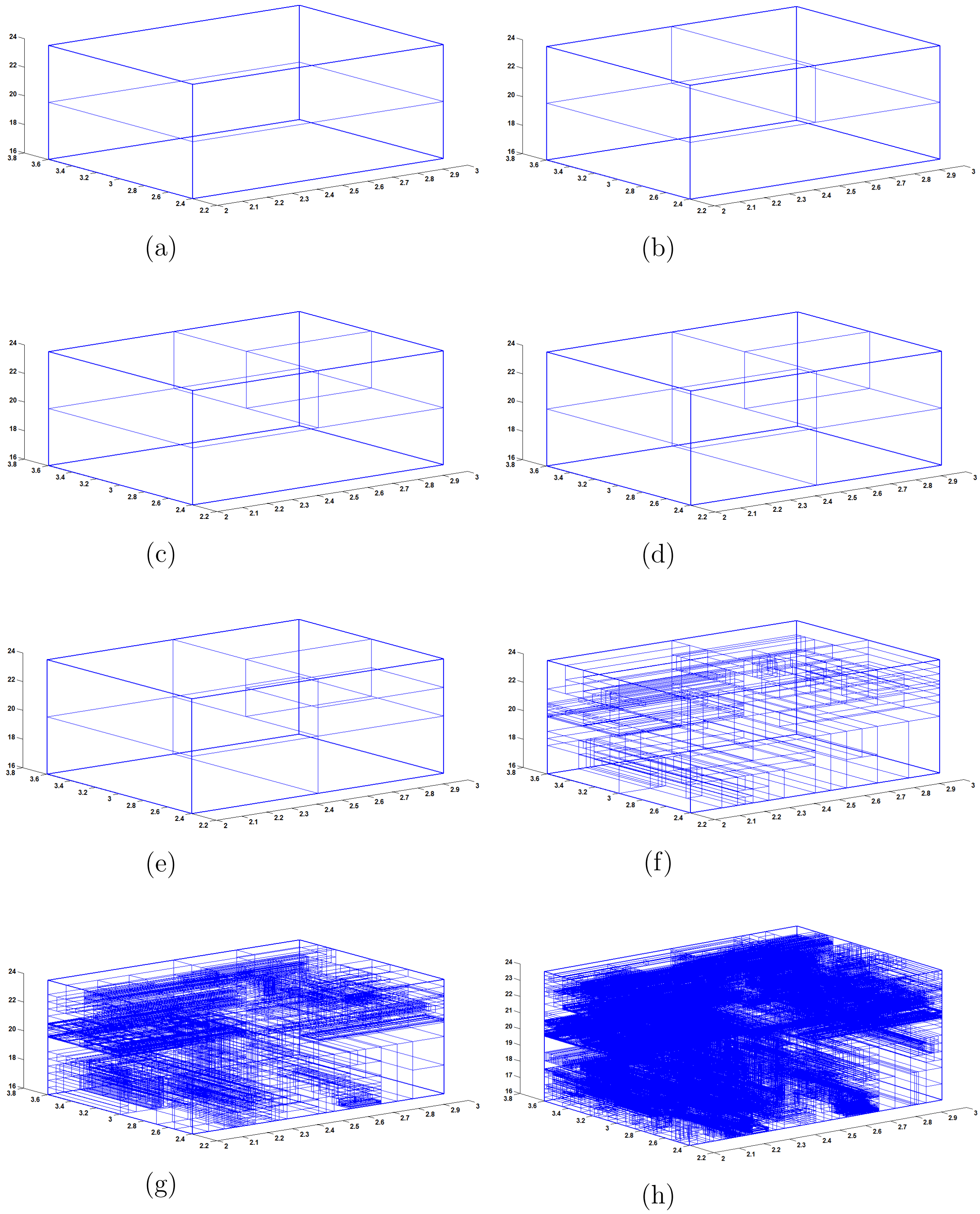



Figure 4.2: Partitioning of support with: (a) 2 cells, (b) 3 cells, (c) 4 cells, (d) 5 cells, (e) 6 cells, (f) 100 cells, (g) 1000 cells, (h) 15000 cells.

## 4.2
## Solution Algorithm with Worst-case Partition Refinement (SAWPR)

In this chapter a more efficient algorithm to partition the support of $\xi$ is presented.

The idea is reducing at each iteration the bounds difference gap splitting the cell at the dimension to make worst the recourse action, using the more conservative previous optimal solution (upper bound optimal solution $\bar{x}^*$).

## 4.3
## How to Choose the Dimension

As mentioned above, the dimension that is going to be split corresponds with the direction in which the recourse action becomes worst, i.e., the dimension that is free to move to the worst case to attained the maximum of the function $Q(\bar{x}^*, \xi)$, for a given $\bar{x}^*$. This adaptation of the robustness is inspired by (Bertsimas and Sim, 2004); according to that, for a given $I_k$, we have to solve the following linear programming problem:

$$\begin{aligned}
\max_{\xi \in I_k, z} \quad & Q(\bar{x}^*, \xi) \\
\text{s.t.} \quad & \bar{\xi}_i^{(k)} - z_i(\bar{\xi}_i^{(k)} - \xi_i^{(k,-)}) \leq \xi_i \leq \bar{\xi}_i^{(k)} + z_i(\xi_i^{(k,+)} - \bar{\xi}_i^{(k)}) \quad \forall i, \\
& \sum_{i=1}^{d} z_i \leq 1, \\
& 0 \leq z_i \leq 1 \quad \forall i.
\end{aligned} \tag{4-1}$$

Let

$$\mathscr{R}(I_k) = \left\{ \xi \in \Xi \;\middle|\; \begin{array}{c} \bar{\xi}_i^{(k)} - z_i(\bar{\xi}_i^{(k)} - \xi_i^{(k,-)}) \leq \xi_i \leq \bar{\xi}_i^{(k)} + z_i(\xi_i^{(k,+)} - \bar{\xi}_i^{(k)}) \quad \forall i, \\ \sum_{i=1}^{d} z_i \leq 1, \\ 0 \leq z_i \leq 1 \quad \forall i \end{array} \right\}.$$

Note that the optimal value $\vartheta^*(z)$ of the linear programming problem

$$\begin{aligned}\vartheta^*(z) = &\operatorname*{Max}_{\xi \in I_k} \quad Q(\bar{x}^*, \xi)\\ &\text{s.t.} \quad \bar{\xi}_i^{(k)} - z_i(\bar{\xi}_i^{(k)} - \xi_i^{(k,-)}) \leq \xi_i \leq \bar{\xi}_i^{(k)} + z_i(\xi_i^{(k,+)} - \bar{\xi}_i^{(k)}) \quad \forall i,\end{aligned}$$

for a fixed $z \in [0,1]$ without the restriction $\sum_{i=1}^d z_i \leq 1$, increase if $z$ increase in any dimension $i \in \{1,\ldots,d\}$, i.e., if $z' \in [0,1]$ is such that $z'_s \geq z_s$ for some $s \in \{1,\ldots,d\}$, then

$$\begin{aligned}&\{\xi \in \Xi \,|\, \bar{\xi}_i^{(k)} - z_i(\bar{\xi}_i^{(k)} - \xi_i^{(k,-)}) \leq \xi_i \leq \bar{\xi}_i^{(k)} + z_i(\xi_i^{(k,+)} - \bar{\xi}_i^{(k)}) \quad \forall i\} \subseteq\\ &\qquad \{\xi \in \Xi \,|\, \bar{\xi}_i^{(k)} - z'_i(\bar{\xi}_i^{(k)} - \xi_i^{(k,-)}) \leq \xi_i \leq \bar{\xi}_i^{(k)} + z'_i(\xi_i^{(k,+)} - \bar{\xi}_i^{(k)}) \quad \forall i\},\end{aligned}$$

and therefore $\vartheta^*(z) \leq \vartheta^*(z')$.

That means that the optimal solution for (4-1) is searched increasing $z$ in all directions, but with the restriction $\sum_{i=1}^d z_i \leq 1$, we impose that this searching must be made it in just one of them, on which the uncertain variable is free to move to the worst case.

To solve (4-1) easier, we can see the following result:

Let

$$V(I_k) = \left\{ \xi \in I_k \;\middle|\; \begin{array}{c} \xi_i \in \{\xi_i^{(k,-)}, \xi_i^{(k,+)}\} \wedge \xi_s = \bar{\xi}_s^{(k)} \quad \forall s \in \{1,\ldots,d\} \setminus \{i\}, \\ \forall i \in \{1,\ldots,d\} \end{array} \right\}.$$

be the set of points for which just one dimension includes the extreme values for the uncertain variable (worst case), and the reaminder hold in the mean scenario (center of the cell).

**Proposition 4.3.1.** *The set of restrictions for (4-1) is equal to the convex hull of the set $V(I_k)$, i.e. $\mathscr{R}(I_k) = conv V(I_k)$.*

***Proof .***

$$
\begin{aligned}
&\xi \in conv V(I_k) \\
&\Longleftrightarrow \xi = \sum_{j=1}^{|V(I_k)|} \theta_j \cdot \xi^j \quad with \quad 0 \le \theta_j \le 1\, \forall j, \sum_{j=1}^{|V(I_k)|} \theta_j = 1, \quad and \quad \xi^j \in V(I_k)\, \forall j \\
&\Longleftrightarrow \xi_i = \sum_{j=1}^{|V(I_k)|} \theta_j \cdot \xi_i^j, \quad \forall i \\
&\Longleftrightarrow \bar{\xi}_i^{(k)} \cdot \sum_{j \in J_i} \theta_j + \xi_i^{(k,-)} \cdot \sum_{j \in J_i^-} \theta_j + \xi_i^{(k,+)} \cdot \sum_{j \in J_i^+} \theta_j, \quad \forall i
\end{aligned}
$$

*where* $J_i \cup J_i^- \cup J_i^+ = \{1, 2, \ldots, |V(I_k)|\}$.
*We have that*

$$\sum_{j \in J_i} \theta_j + \sum_{j \in J_i^-} \theta_j + \sum_{j \in J_i^+} \theta_j = \sum_{j=1}^{|V(I_k)|} \theta_j = 1.$$

*Let*

$$z_i = 1 - \sum_{j \in J_i} \theta_j = \sum_{j \in J_i^-} \theta_j + \sum_{j \in J_i^+} \theta_j \ge 0.$$

*so*

$$\Longleftrightarrow \begin{cases} \xi_i & \ge \bar{\xi}_i^{(k)} \cdot \underbrace{\sum_{j \in J_i} \theta_j}_{=1-z_i} + \left( \underbrace{\sum_{j \in J_i^-} \theta_j + \sum_{j \in J_i^+} \theta_j}_{=z_i} \right) \cdot \xi_i^{(k,-)} \\ & \ge \bar{\xi}_i^{(k)}(1 - z_i) + z_i \xi_i^{(k,-)} \\ & = \bar{\xi}_i^{(k)} - z_i(\bar{\xi}_i^{(k)} - \xi_i^{(k,-)}) \end{cases}$$

$$\Longleftrightarrow \begin{cases} \xi_i & \le \bar{\xi}_i^{(k)} \cdot \sum_{j \in J_i} \theta_j + \left( \sum_{j \in J_i^-} \theta_j + \sum_{j \in J_i^+} \theta_j \right) \cdot \xi_i^{(k,+)} \\ & \le \bar{\xi}_i^{(k)}(1 - z_i) + z_i \xi_i^{(k,+)} \\ & = \bar{\xi}_i^{(k)} + z_i(\xi_i^{(k,+)} - \bar{\xi}_i^{(k)}) \end{cases}$$

*Therefore*

$$\bar{\xi}_i^{(k)} - z_i(\bar{\xi}_i^{(k)} - \xi_i^{(k,-)}) \le \xi_i \le \bar{\xi}_i^{(k)} + z_i(\xi_i^{(k,+)} - \bar{\xi}_i^{(k)}) \quad \forall i,$$

*Also,*

$$\sum_{i=1}^{d} \left( \sum_{j \in J_i} \theta_j \right) = d - 1,$$

*and therefore*

$$\sum_{i=1}^{d} z_i = d - \sum_{i=1}^{d} \left( \sum_{j \in J_i} \theta_j \right) = 1$$

$\Longleftrightarrow \xi \in \mathscr{R}(I_k)$. $\square$

As the set of extreme points of $\operatorname{conv} V(I_k)$ is the set $V(I_k)$ itself, and $Q(\bar{x}^*, \xi)$ is a convex function, we know that (4-1) is equivalence to

$$\underset{\xi \in V(I_k)}{\operatorname{Max}} \; Q(\bar{x}^*, \xi), \tag{4-2}$$

which is computationally easier to solve.

## 4.4
## How to Choose the Cell

Since we want to reduce the gap between both bounds at each iteration, we will choose the cell where the distance between the upper bound and the lower bound is the biggest. So, we have to solve:

$$k^* \in \underset{k}{\operatorname{Argmax}} \left( \underset{\xi \in E(I_k)}{\operatorname{Max}} \; Q(\bar{x}^*, \xi) - Q(\bar{x}^*, \mathbb{E}[\xi | \xi \in I_k]) \right) \tag{4-3}$$

Choosing the cell in this way, we are trying to equalize the convergence velocity for both bounds.

**Algorithm 3** SAWPR

**Require:** $\bar{x}^*$ and $\mathcal{I} = \{I_k\}_{k=1}^{n}$
**Ensure:** $\mathcal{J} = \{J_{k'}\}_{k'=1}^{n+1}$ such that $\mathcal{J}$ refines $\mathcal{I}$

**if** n = 1 **then**
    Put

$$\begin{aligned} \xi^* \in \operatorname*{Argmax}_{\xi \in \Xi,\, z} \quad & Q(\bar{x}^*, \xi) \\ \text{s.t.} \quad & \bar{\xi}_i - z_i(\bar{\xi}_i - \xi_i^-) \leq \xi_i \leq \bar{\xi}_i + z_i(\xi_i^+ - \bar{\xi}_i) \quad \forall i \\ & \sum_{i=1}^{d} z_i \leq 1 \\ & 0 \leq z_i \leq 1 \quad \forall i. \end{aligned}$$

    Let $i^* \in \{1, \ldots, d\}$ be such that $\xi^*_{i^*} \in \{\xi^-_{i^*}, \xi^+_{i^*}\}$
    Put $J_1 = [\xi^-_{i^*}, \bar{\xi}_{i^*}] \times \prod_{t \in \{1,2,\ldots,d\} \setminus \{i^*\}} [\xi^-_t, \xi^+_t]$ and $J_2 = [\bar{\xi}_{i^*}, \xi^+_{i^*}] \times \prod_{t \in \{1,2,\ldots,d\} \setminus \{i^*\}} [\xi^-_t, \xi^+_t]$
    $\mathcal{J} = \{J_{k'}\}_{k'=1}^{2}$
**else**
    Find $k^* \in \operatorname*{Argmax}_{k} \left( \operatorname*{Max}_{\xi \in E(I_k)} Q(\bar{x}^*, \xi) - Q(\bar{x}^*, E[\xi | \xi \in I_k]) \right)$
    Put

$$\begin{aligned} \xi^* \in \operatorname*{Argmax}_{\xi \in I_{k^*},\, z} \quad & Q(\bar{x}^*, \xi) \\ \text{s.t.} \quad & \bar{\xi}_i^{(k^*)} - z_i(\bar{\xi}_i^{(k^*)} - \xi_i^{(k^*,-)}) \leq \xi_i \leq \bar{\xi}_i^{(k^*)} + z_i(\xi_i^{(k^*,+)} - \bar{\xi}_i^{(k^*)}) \quad \forall i \\ & \sum_{i=1}^{d} z_i \leq 1 \\ & 0 \leq z_i \leq 1 \quad \forall i. \end{aligned}$$

    Let $i^* \in \{1, \ldots, d\}$ be such that $\xi^*_{i^*} \in \{\xi^-_{i^*}, \xi^+_{i^*}\}$
    Put $k' \Longleftarrow k \; \forall k \neq k^*$, $J_{k^*} \Longleftarrow [\xi_{i^*}^{(k^*,-)}, \bar{\xi}_{i^*}^{(k^*)}] \times \prod_{t \in \{1,2,\ldots,d\} \setminus \{i^*\}} [\xi_t^{(k^*,-)}, \xi_t^{(k^*,+)}]$ and
    $J_{n+1} \Longleftarrow [\bar{\xi}_{i^*}^{(k^*)}, \xi_{i^*}^{(k^*,+)}] \times \prod_{t \in \{1,2,\ldots,d\} \setminus \{i^*\}} [\xi_t^{(k^*,-)}, \xi_t^{(k^*,+)}]$
    $\mathcal{J} = \{J_{k'}\}_{k'=1}^{n+1}$
**end if**

## 4.5 Graphyc Explanation of the SAWPR

Since the Farmer's problem is dimensionally representable, we will take this problem as an example of two-stage stochastic linear programming problem with complete recourse, to explain graphically the working of the SAWPR algorithm.

At the first stage we have to choose the dimension that is going to be split to get the two initial cells of the first partition. To do that, it is necessary find the optimal solution of the upper bound problem (5-5) to get the initial allocation $\bar{x}_0^*$ requiered to solve:

$$\begin{aligned} \xi^* \in \underset{\xi\in\Xi,\, z}{\operatorname{Argmax}} \quad & Q(\bar{x}_0^*, \xi) \\ \text{s.t.} \quad & \bar{\xi}_i - z_i(\bar{\xi}_i - \xi_i^-) \leq \xi_i \leq \bar{\xi}_i + z_i(\xi_i^+ - \bar{\xi}_i) \quad \forall i \\ & \sum_{i\in\{T,M,B\}} z_i \leq 1 \\ & 0 \leq z_i \leq 1 \quad \forall i \in \{T, M, B\}. \end{aligned}$$

As was mentioned above that problem is equivalent to:

$$\xi^* \in \underset{\xi\in V(I_0)}{\operatorname{Argmax}} Q(\bar{x}_0^*, \xi),$$

where

$$V(I_0) = \left\{ \xi \in \Xi \;\middle|\; \begin{array}{c} \xi_i \in \{\xi_i^-, \xi_i^+\} \wedge \xi_s = \bar{\xi}_s \quad \forall s \in \{T, M, B\} \setminus \{i\}, \\ \forall i \in \{T, M, B\} \end{array} \right\},$$

since $V(I_0)$, is itself the set of extreme points of the convex hull $\operatorname{conv} V(I_0)$.

After that, according to **Algorithm 3**, we have to identify the dimension that was free to move to the worst case (extreme value) in $\xi^*$, which was mathematically expressed by:

$$i^* \in \{T, M, B\} \quad \text{s.t.} \quad \xi^*_{i^*} \in \{\xi^-_{i^*}, \xi^+_{i^*}\}.$$

Assuming that $\xi_T \sim U([2,3]$, $\xi_M \sim U([2.4, 3.6])$ and $\xi_B \sim U([16, 24])$, the support of $\xi$ and the convex hull (shaded region) of $V(I_0)$, are represented in Figure 4.3.

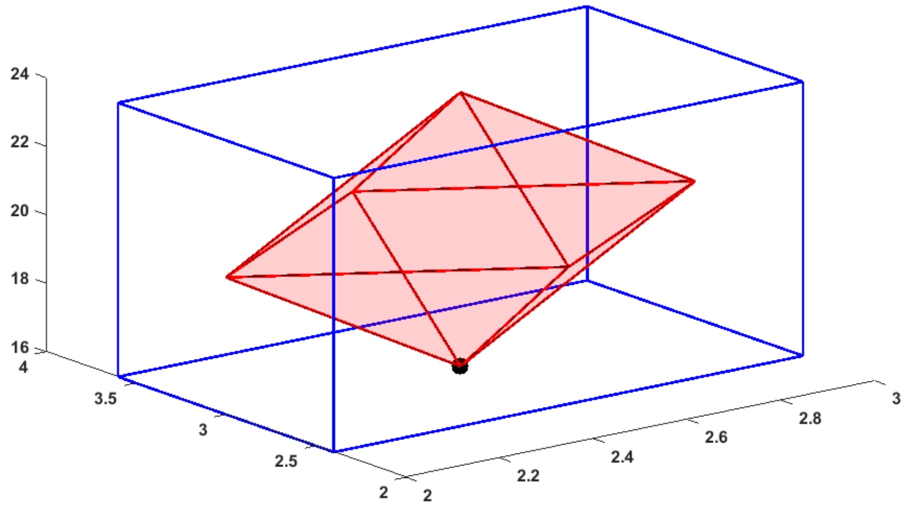



Figure 4.3: Convex hull (shaded region) for the set $V(I_0)$, where $I_0 = \Xi$ is the original support of $\xi$. The black point is the optimal solution for the maximization problem of $Q(\bar{x}^*_0, \xi)$ with $\xi \in \text{conv}V(I_0)$.

Evaluating the recourse function $Q(\bar{x}^*_0, \xi)$ for the initial allocation $\bar{x}^*_0$ at each point of the set $V(I_0)$, we can find the vertex where the maximum of $Q(\bar{x}^*_0, \xi)$ is attained, and with this, identify the dimension where the uncertain variable leave the center of the cell.

According to the Figure 4.3, the maximum was attained at the point $\xi^* = (\bar{\xi}_T, \bar{\xi}_M, \xi_B^-)^t = (2.5, 3, 16)^t$, which means that the uncertain variable move to the worst case for the third dimension (beet dimension). So the initial cut is made in the mean of $\xi_B$ with support equal to $[\xi_B^-, \xi_B^+] = [16, 24]$, as shown in the Figure 4.4.

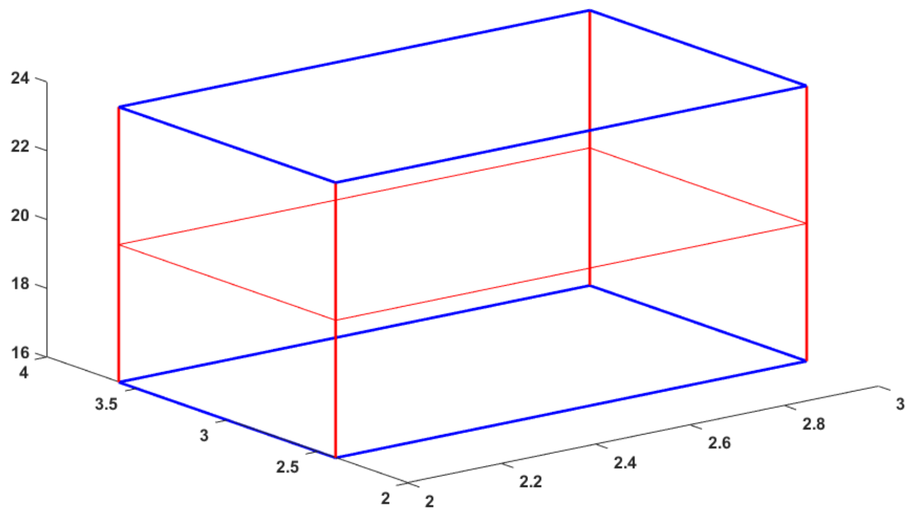



Figure 4.4: Initial cut in the mean $\bar{\xi}_B = 20$.

The red edges indicate the dimension that was chosen to be splitted and the red plane represented the cut.

With the two initials cells:

$$
\begin{aligned}
I_1 &= [\xi_T^-, \xi_T^+] \times [\xi_M^-, \xi_M^+] \times [\xi_B^-, \bar{\xi}_B] = [2,3] \times [2.4, 3.6] \times [16, 20],\\
I_2 &= [\xi_T^-, \xi_T^+] \times [\xi_M^-, \xi_M^+] \times [\bar{\xi}_B, \xi_B^+] = [2,3] \times [2.4, 3.6] \times [20, 24],
\end{aligned}
$$

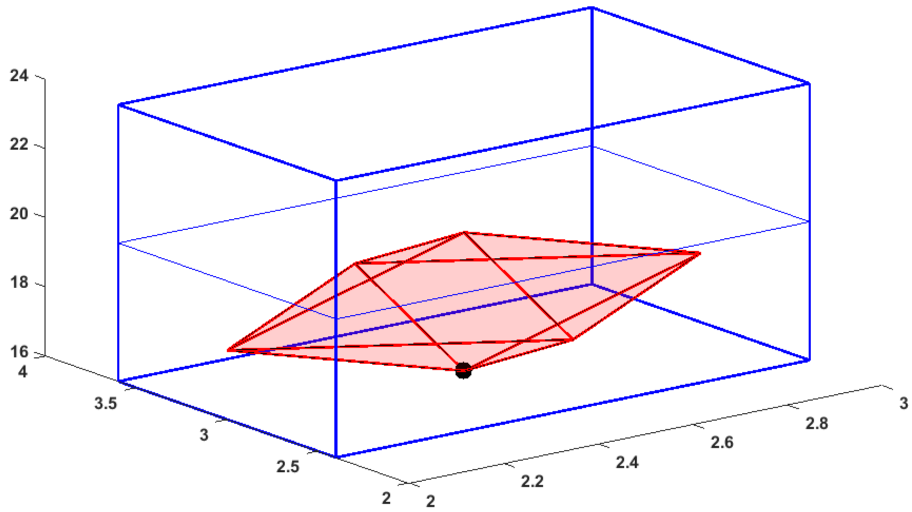
we solve the upper (5-5) and lower (5-9) bound problem.

Considering the last optimal solution for the upper bound problem $\bar{x}^*$ as a new estimated allocation, we have to choose which of the two existing cells is going to be split to refine the actual partition, solving the problem mentioned in (4-3). The goal is identifying the cell where the difference gap of the lower and upper approximation of the expected value of the recourse function is bigger. Mathematically this was expressed by:

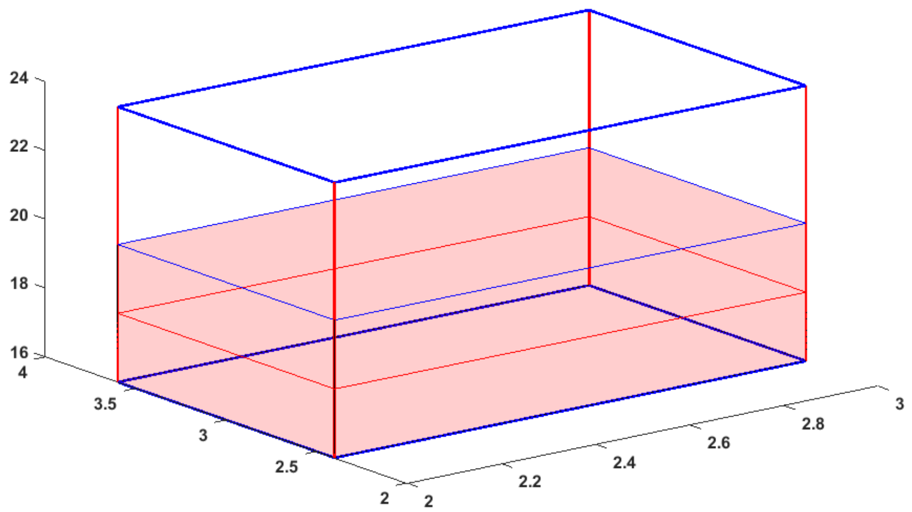
$$
k^* \in \operatorname*{Argmax}_k \left( \operatorname*{Max}_{\xi \in E(I_k)} Q(\bar{x}^*, \xi) - Q(\bar{x}^*, \mathbb{E}[\xi | \xi \in I_k]) \right).
$$

For this initial partition, the chosen cell was $I_1 = [2,3]\times[2.4,3.6]\times[14,20]$.

At this stage we have to come back to the initial problem about choose the dimension at which the cut will be made, but that is done in the same way considering this time the optimal cell $I_1$ instead of the original support $\Xi$.




(a)




(b)

Figure 4.5: (a) Convex hull (shaded region) for the set $V(I_1)$, where $I_1 = [2,3] \times [2.4,3.6] \times [14,20]$. The black point is the optimal solution for the maximization problem of $Q(\bar{x}^*,\xi)$ with $\xi \in \text{conv}V(I_1)$. (b) Choose of the cell $I_1$, as the optimal cell.

Doing this, the chosen dimension is the third one (beet dimension) again (see Figure 4.5), and cutting in the mean of $\xi_B$ in the cell $I_1$ with support $[16, 20]$, we obtain the new partition with the three cells:

$$
\begin{aligned}
I_1 &= [2,3] \times [2.4, 3.6] \times [16, 18],\\
I_2 &= [2,3] \times [2.4, 3.6] \times [18, 20],\\
I_3 &= [2,3] \times [2.4, 3.6] \times [20, 24].
\end{aligned}
$$

After this we have to solve the upper (5-5) and lower (5-9) bound problem, respectively, to get an accurate aproximation of the true problem (2-24).

At the next stage we have to choose the optimal cell to refine the actual partition $\{I_1, I_2, I_3\}$. Solving (4-3), we find this time that the cell where the difference between the lower and upper aproximation of the expected value of the recourse action is bigger, is the cell $I_1 = [2,3] \times [2.4, 3.6] \times [16, 18]$, represented by the shaded region in the Figure 4.6.



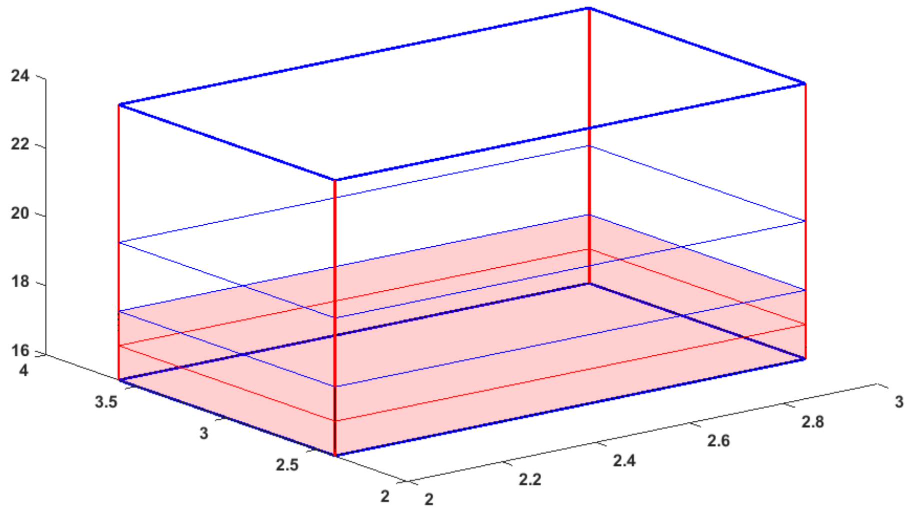



Figure 4.6: Choose of the cell $I_1 = [2,3] \times [2.4, 3.6] \times [16, 18]$, as the optimal cell.

For the cell $I_1 = [2,3] \times [2.4, 3.6] \times [16, 18]$, the optimal solution of the problem (4-1) pointed again the third dimension (beet dimension) as the optimal direction to do the cut. Doing this, the newest partition is composed by the four cells:

$$I_1 = [2,3] \times [2.4, 3.6] \times [16, 17],$$
$$I_2 = [2,3] \times [2.4, 3.6] \times [17, 18],$$
$$I_3 = [2,3] \times [2.4, 3.6] \times [18, 20],$$
$$I_4 = [2,3] \times [2.4, 3.6] \times [20, 24].$$

This sequencial problems are solved until reach a tolerance level for the difference between the lower and upper optimal approximate solutions, established at the begining as an stop criteria for the algorithm.

In the next figure is presented the partitioning of support for different iterations of the algorithm (partition size).

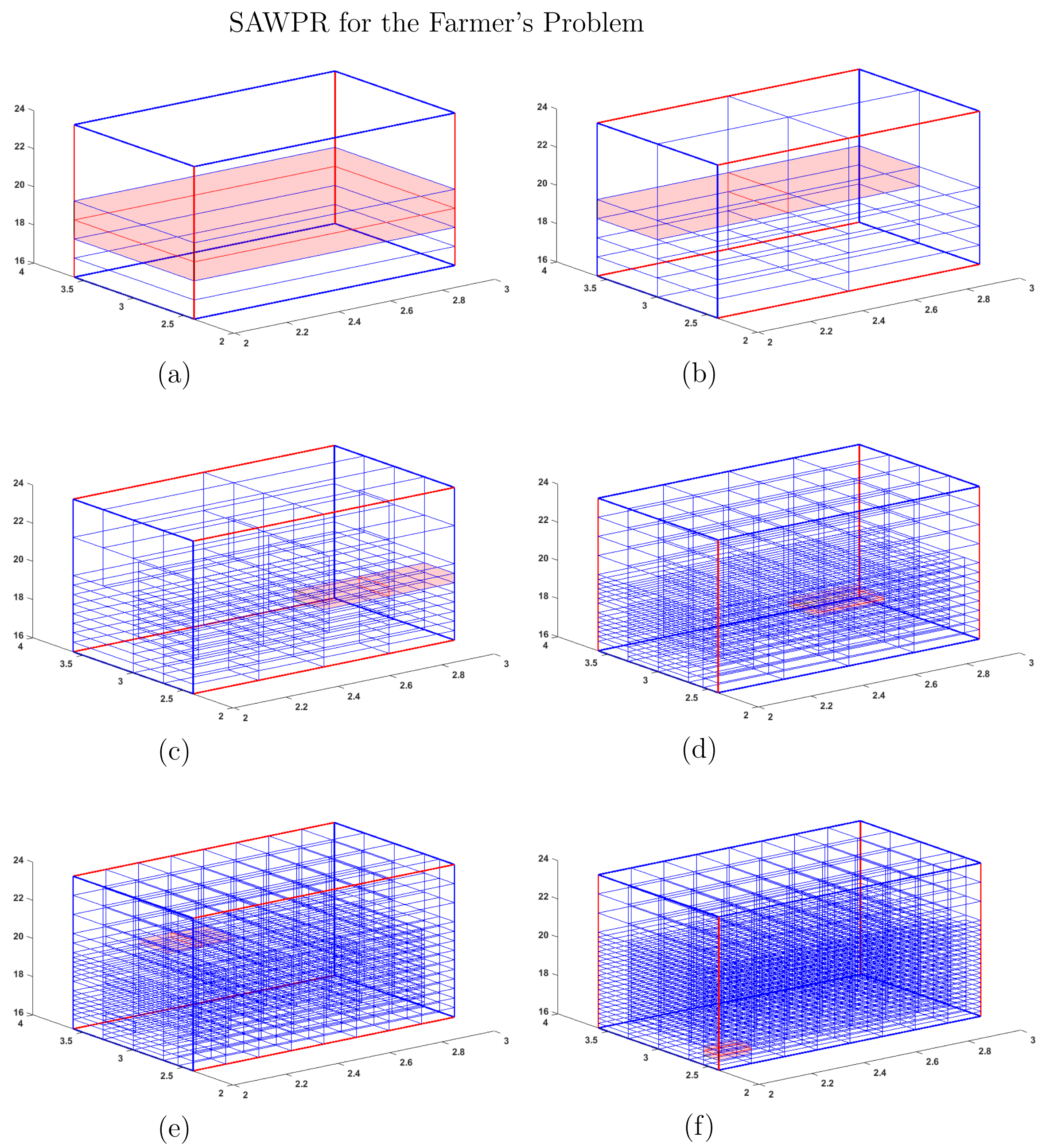



Figure 4.7: Partition of the support with: (a) 4 cells, (b) 20 cells, (c) 100 cells, (d) 300 cells, (e) 500 cells, (f) 900 cells.

# 5
# Case Study

## 5.1
## Bounding the Uncertainty

We clarify that we consider problems with complete recourse like the Farmer's problem, for which the following treatment is valid.

We saw that discretizing the expected value operator we can obtain bounds for the optimal solution of the true problem. One bound is obtained considering the weighted sum of the worst values of $Q(x,\xi)$ for the case of minimizing problem, and the other one is obtained considering the weighted sum of $Q(x,\xi)$ evaluated in the center of each cell.

The aim is verifing the convergence of the partition-based method considering the first conditional moment.

According to this, the true problem (2-24) can be write as:

$$\begin{aligned} \underset{x}{\text{Min}} \quad & c^t x + \sum_{k=1}^{n} \mathbb{E}_\xi[Q(x,\xi)|\xi \in I_k] P(\xi \in I_k) \\ \text{s.t.} \quad & x \in X. \end{aligned} \tag{5-1}$$

Additionally, we assume for simplicity that $\{\xi_T, \xi_M, \xi_B\}$ is a set of independent random variables with uniform distribution, such that $\xi_i \sim U([\xi_i^-, \xi_i^+])$, for all $i \in \{T, M, B\}$. Therefore the probability of the random variable $\xi$ belongs to cell $I_k$, is given by:

$$P(\xi \in I_k) = P(\xi_T \in [\xi_T^{(k,-)}, \xi_T^{(k,+)}]) \cdot P(\xi_M \in [\xi_M^{(k,-)}, \xi_M^{(k,+)}]) \cdot P(\xi_B \in [\xi_B^{(k,-)}, \xi_B^{(k,+)}]),$$

where

$$\begin{aligned} P(\xi_i \in [\xi_i^{(k,-)}, \xi_i^{(k,+)}]) &= P(\xi_i \le \xi_i^{(k,+)}) - P(\xi_i \le \xi_i^{(k,-)}) \\ &= \frac{\xi_i^{(k,+)} - \xi_i^{(k,-)}}{\xi_i^+ - \xi_i^-}, \end{aligned}$$

for all $i \in \{T, M, B\}$.

## 5.2 Upper Bound

We have that

$$\begin{cases} \underset{x}{\text{Min}} & c^t x + \sum_{k=1}^{n} \mathbb{E}_\xi[Q(x,\xi)|\xi \in I_k] P(\xi \in I_k) \\ \text{s.t.} & x \in X. \end{cases} \leq$$

$$\begin{cases} \underset{x}{\text{Min}} & c^t x + \sum_{k=1}^{n} \underset{\xi \in I_k}{Max}\, Q(x,\xi) \cdot P(\xi \in I_k) \\ \text{s.t.} & x \in X. \end{cases}$$

So, the upper bound for the optimal solution of Farmer's problem is given by:

$$\begin{aligned} \underset{x}{\text{Min}} \quad & 150x_T + 230x_M + 260x_B + \sum_{k=1}^{n} \max_{\xi \in I_k} Q(x,\xi) \cdot P(\xi \in I_k) \\ \text{s.t.} \quad & x_T + x_M + x_B \leq 500, \\ & x_T, x_M, x_B \geq 0, \end{aligned} \tag{5-2}$$

where $Q(x,\xi)$, is defined as in (2-25).

We know that

$$\begin{aligned} \underset{\xi \in I_k}{\text{Max}}\, Q(x,\xi) &= \underset{\xi \in E(I_k)}{\text{Max}}\, Q(x,\xi) \\ &= \max\{Q(x,\xi^{(k,s)}) | s \in \{-,+\}^3\}, \end{aligned}$$

where $\xi^{(k,s)}$, is the value of $\xi$ on the vertex $s$ of the cell $k$, and $x$ is fixed.

To clarify the notation, $\{-,+\}^3$ it denotes the set of coordenates for the vertex of cell $k$, that are formed by the intervals ends of each dimension.

$$\{-,+\}^3 = \{-,+\} \times \{-,+\} \times \{-,+\} = \begin{Bmatrix} (-,-,-),(+,-,-),(+,+,-), \\ (-,+,-),(-,-,+),(+,-,+), \\ (+,+,+),(-,+,+) \end{Bmatrix},$$

thus, for example, if $s = (-,-,+)$, then $\xi^{(k,s)} = (\xi_T^{(k,-)}, \xi_M^{(k,-)}, \xi_B^{(k,+)})$.

Retaking the above, (3-29) can be expressed by the following linear programming problem:

$$\begin{aligned} \underset{\theta_k}{\text{Min}} \quad & \theta_k \\ \text{s.t.} \quad & \theta_k \geq Q(x, \xi^{(k,s)}) \quad \forall s \in \{-,+\}^3, \end{aligned} \tag{5-3}$$

where

$$\begin{aligned} Q(x,\xi^{(k,s)}) = \underset{y,w}{\text{Min}} \quad & 238y_T^{(k,s)} + 210y_M^{(k,s)} - 170w_T^{(k,s)} - 150w_M^{(k,s)} - 36w_{B_1}^{(k,s)} - 10w_{B_2}^{(k,s)} \\ \text{s.t.} \quad & \xi_T^{(k,s)} x_T + y_T^{(k,s)} - w_T^{(k,s)} \geq 200, \\ & \xi_M^{(k,s)} x_M + y_M^{(k,s)} - w_M^{(k,s)} \geq 240, \\ & \xi_B^{(k,s)} x_B - w_{B_1}^{(k,s)} - w_{B_2}^{(k,s)} \geq 0, \\ & - w_{B_1}^{(k,s)} \geq -6000, \\ & y_T^{(k,s)}, y_M^{(k,s)} \geq 0, \quad w_T^{(k,s)}, w_M^{(k,s)}, w_{B_1}^{(k,s)}, w_{B_2}^{(k,s)} \geq 0, \ \textit{with } s \in \{-,+\}^3. \end{aligned}$$

Let $\theta_k^*$ be the optimal value of the problem (5-3), so:

$$\theta_k^* = Q(x, \xi^{(k,s^*)}) = \max\{Q(x,\xi^{(k,s)}) | s \in \{-,+\}^3\},$$

where $s^*$ denote the vertex on which the worst value of $Q(x,\xi)$ inside the cell $k$ is.

Attending this, (5-3) is equivalent to the following problem:

$$\begin{aligned} \underset{\theta,y,w}{\text{Min}} \quad & \theta_k \\ \text{s.t.} \quad & \theta_k \geq 238y_T^{(k,s)} + 210y_M^{(k,s)} - 170w_T^{(k,s)} - 150w_M^{(k,s)} - 36w_{B_1}^{(k,s)} - 10w_{B_2}^{(k,s)} \quad \forall s \in \{-,+\}^3, \\ & \xi_T^{(k,s)} x_T + y_T^{(k,s)} - w_T^{(k,s)} \geq 200 \quad \forall s \in \{-,+\}^3, \\ & \xi_M^{(k,s)} x_M + y_M^{(k,s)} - w_M^{(k,s)} \geq 240 \quad \forall s \in \{-,+\}^3, \\ & \xi_B^{(k,s)} x_B - w_{B_1}^{(k,s)} - w_{B_2}^{(k,s)} \geq 0 \quad \forall s \in \{-,+\}^3, \\ & - w_{B_1}^{(k,s)} \geq -6000 \quad \forall s \in \{-,+\}^3, \\ & y_T^{(k,s)}, y_M^{(k,s)} \geq 0, \quad w_T^{(k,s)}, w_M^{(k,s)}, w_{B_1}^{(k,s)}, w_{B_2}^{(k,s)} \geq 0 \quad \forall s \in \{-,+\}^3. \end{aligned} \tag{5-4}$$

For an explanation more detailed about this fact see Indeed A.6

By other hand, is fact that each coordenate $\theta_k^*$ of $\theta^* = (\theta_k^*)_{k=1}^n$, that is the minimizer for the average $\sum_{k=1}^n \theta_k \cdot P(\xi \in I_k)$, is the optimal value of the problem (5-4), for all $k = 1, \ldots, n$ (see A.7).

Considering the above arguments, the problem (5-2) is equivalent to the following linear programing problem:

$$
\begin{aligned}
\min_{x,y,w,\theta} \quad & 150x_T + 230x_M + 260x_B + \sum_{k=1}^{n} \theta_k \cdot P(\xi \in [\xi^{(k,-)}, \xi^{(k,+)}]) \\
\text{s.t.} \quad & -(x_T + x_M + x_B) \geq -500, \\
& \theta_k \geq 238y_T^{(k,s)} + 210y_M^{(k,s)} - 170w_T^{(k,s)} - 150w_M^{(k,s)} - 36w_{B_1}^{(k,s)} - 10w_{B_2}^{(k,s)} \quad \forall s \in \{-,+\}^3, \\
& \xi_T^{(k,s)} x_T + y_T^{(k,s)} - w_T^{(k,s)} \geq 200 \quad \forall s \in \{-,+\}^3, \\
& \xi_M^{(k,s)} x_M + y_M^{(k,s)} - w_M^{(k,s)} \geq 240 \quad \forall s \in \{-,+\}^3, \\
& \xi_B^{(k,s)} x_B - w_{B_1}^{(k,s)} - w_{B_2}^{(k,s)} \geq 0 \quad \forall s \in \{-,+\}^3, \\
& -w_{B_1}^{(k,s)} \geq -6000 \quad \forall s \in \{-,+\}^3, \\
& y_T^{(k,s)}, y_M^{(k,s)} \geq 0, \quad w_T^{(k,s)}, w_M^{(k,s)}, w_{B_1}^{(k,s)}, w_{B_2}^{(k,s)} \geq 0 \quad \forall s \in \{-,+\}^3, \\
& x_T, x_M, x_B \geq 0,
\end{aligned} \tag{5-5}
$$

which optimal value represents the upper bound of the theoretical value for the Farmer's problem.

## 5.3 Lower Bound

We saw that $Q(x,\xi)$ defined as in (2-25), is a convex function. On other hand, by Jensen inequality (see A.4), we have that

$$\mathbb{E}[Q(x,\xi)|\xi \in I_k] \leq Q(x, \mathbb{E}[\xi|\xi \in I_k]) \quad \forall k, \tag{5-6}$$

so

$$
\begin{cases}
\min\limits_{x} & c^t x + \sum\limits_{k=1}^{n} Q(x, \mathbb{E}[\xi|\xi \in I_k]) \cdot P(\xi \in I_k) \\
\text{s.t.} & x \in X.
\end{cases}
\leq
\begin{cases}
\min\limits_{x} & c^t x + \sum\limits_{k=1}^{n} \mathbb{E}_\xi[Q(x,\xi)|\xi \in I_k] P(\xi \in I_k) \\
\text{s.t.} & x \in X.
\end{cases}
$$

In that sense, the lower bound for the true problem (2-24) is given by the following linear programming problem:

$$
\begin{aligned}
\underset{x}{\text{Min}} \quad & 150x_T + 230x_M + 260x_B + \sum_{k=1}^{n} Q(x, E[\xi|\xi \in I_k]) \cdot P(\xi \in I_k) \\
\text{s.t.} \quad & x_T + x_M + x_B \leq 500, \\
& x_T, x_M, x_B \geq 0,
\end{aligned}
\tag{5-7}
$$

where

$$
\begin{aligned}
Q(x, E[\xi|\xi \in I_k]) = \\
\underset{y,w}{\text{Min}} \quad & 238y_T^{(k)} + 210y_M^{(k)} - 170w_T^{(k)} - 150w_M^{(k)} - 36w_{B_1}^{(k)} - 10w_{B_2}^{(k)} \\
\text{s. t.} \quad & \bar{\xi}_T^{(k)} x_T + y_T^{(k)} - w_T^{(k)} \geq 200, \\
& \bar{\xi}_M^{(k)} x_M + y_T^{(k)} - w_M^{(k)} \geq 240, \\
& \bar{\xi}_B^{(k)} x_B + w_{B_1}^{(k)} - w_{B_2}^{(k)} \geq 0, \\
& - w_{B_1}^{(k)} \geq -6000 \\
& .
\end{aligned}
\tag{5-8}
$$

Again, under the same arguments, it is possible to solve at the same time the first and the second stage problem, and therefore (5-7) is equivalen to :

$$
\begin{aligned}
\underset{x,y,w}{\text{Min}} \quad & \left\{ \begin{aligned} & 150x_T + 230x_M + 260x_B + \\ & \sum_{k=1}^{n} \left\{ \begin{aligned} & 238y_T^{(k)} + 210y_M^{(k)} - 170w_T^{(k)} \\ & - 150w_M^{(k)} - 36w_{B_1}^{(k)} - 10w_{B_2}^{(k)} \end{aligned} \right\} P(\xi \in I_k) \end{aligned} \right. \\
\text{s.t.} \quad & x_T + x_M + x_B \leq 500, \\
& x_T, x_M, x_B \geq 0, \\
& \bar{\xi}_T^{(k)} x_T + y_T^{(k)} - w_T^{(k)} \geq 200 \quad \forall k = 1, \ldots, n, \\
& \bar{\xi}_M^{(k)} x_M + y_M^{(k)} - w_M^{(k)} \geq 240 \quad \forall k = 1, \ldots, n, \\
& \bar{\xi}_B^{(k)} x_B - w_{B_1}^{(k)} - w_{B_2}^{(k)} \geq 0 \quad \forall k = 1, \ldots, n, \\
& w_{B_1}^{(k)} \leq 6000 \quad \forall k = 1, \ldots, n, \\
& y_T^{(k)}, y_M^{(k)} \geq 0 \quad \forall k = 1, \ldots, n, \\
& w_T^{(k)}, w_M^{(k)}, w_{B_1}^{(k)}, w_{B_2}^{(k)} \geq 0 \quad \forall k = 1, \ldots, n.
\end{aligned}
\tag{5-9}
$$

# 6
# Computational Results

## 6.1
## SAA for the Farmer's Problem

Here are presented the obtained results after running the model (2-26), increasing sufficiently the scenarios number to verify the convergence of the SAA method. The error was estimated by the variance of the solution with 100 replications of the problem at each iteration.

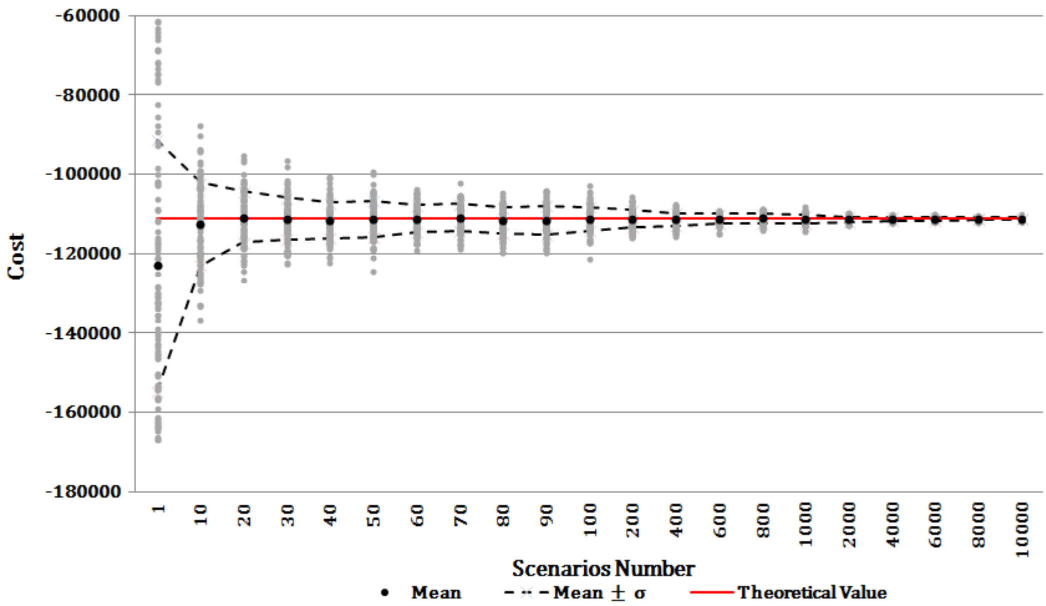



Figure 6.1: SAA estimator for the Framer's Problem

Observing the Figure 6.1 it can be seen that there exist an inverse relation between the scenarios number and the error. The dispersion vanishes when the sample size is sufficiently big to get an accurate approximate solution.

However the computational cost to run the large linear programming problem (2-26) 100 times with a big scenarios number, is very high; that fact leads us to think about if the desired accuracy for a more complex problem with other kind of uncertain data could be unachievable.

Moreover we can verify that the SAA estimator is downwards biased for the cost minimization problem, and that bias decrease monotonically when the number of scenarios increases. So, in that sense, a conservatism solution would be $\mathbb{E}[\hat{\vartheta}_N^*] + \hat{\sigma}(x^*)$, where $\hat{\sigma}^2(x^*)$ denoted the sample variance.

## 6.2 Numerical Results from the SARPR

The models (5-5) and (5-9) were run, and we obtained the following results:

| Partition Size[1] | Upper Bound | Lower Bound |
|---|---|---|
| 1 | -59950.00 | -118600.00 |
| 20 | -70069.31 | -114632.50 |
| 40 | -72496.29 | -114480.63 |
| 60 | -73256.68 | -114477.81 |
| 80 | -73478.08 | -113474.69 |
| 100 | -73822.43 | -113471.87 |
| 200 | -87649.93 | -113466.25 |
| 400 | -90236.27 | -113434.74 |
| 600 | -90767.71 | -112636.99 |
| 800 | -93228.96 | -112258.07 |
| 1000 | -93777.27 | -112255.73 |
| 2000 | -94814.96 | -112019.48 |
| 4000 | -96950.18 | -111876.00 |
| 6000 | -97642.41 | -111608.13 |
| 8000 | -98296.21 | -111538.18 |
| 10000 | -98687.31 | -111522.92 |
| 15000 | -100048.17 | -111511.24 |

Table 6.1: Numerical results.

This data are represented in the Figure 6.2. It can be appreciated the convergence of the partition-based method and the accuracy of its solution.

The Figure 6.3 shows the percentual error. The gray area corresponds to percentage of the true solution that was reached by the partition-based method, and the darker one corresponds to the error.

Comparing both bounds (see Figure 6.4), we note that their convergence velocity is different.

If we analyzed the formulation of bounding problem for the minimization case, it is clear that the conservatism level for the upper bound problem (robustness of the recourse action) is higher than the mean scenario one. In that sense it is expected that the upper bound converges slower than lower one.

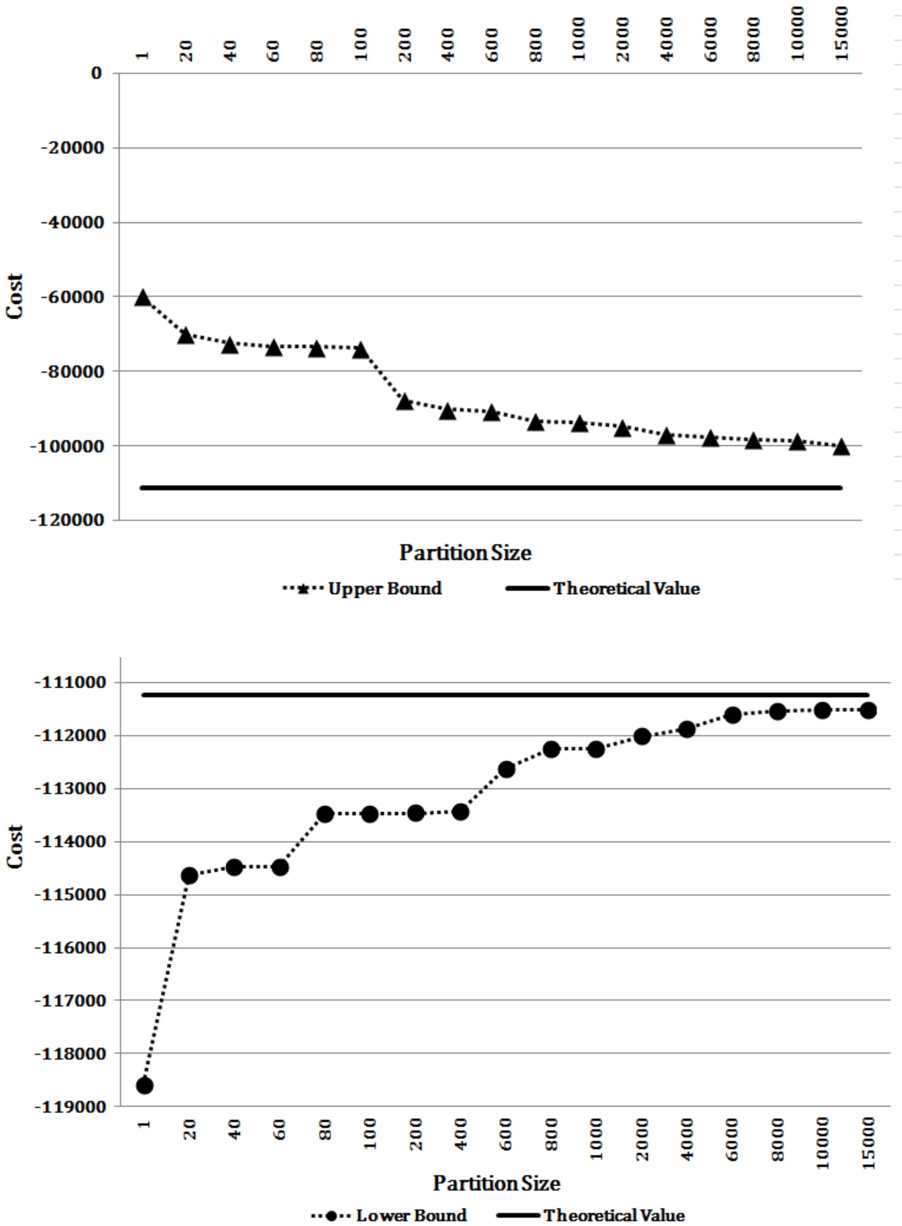
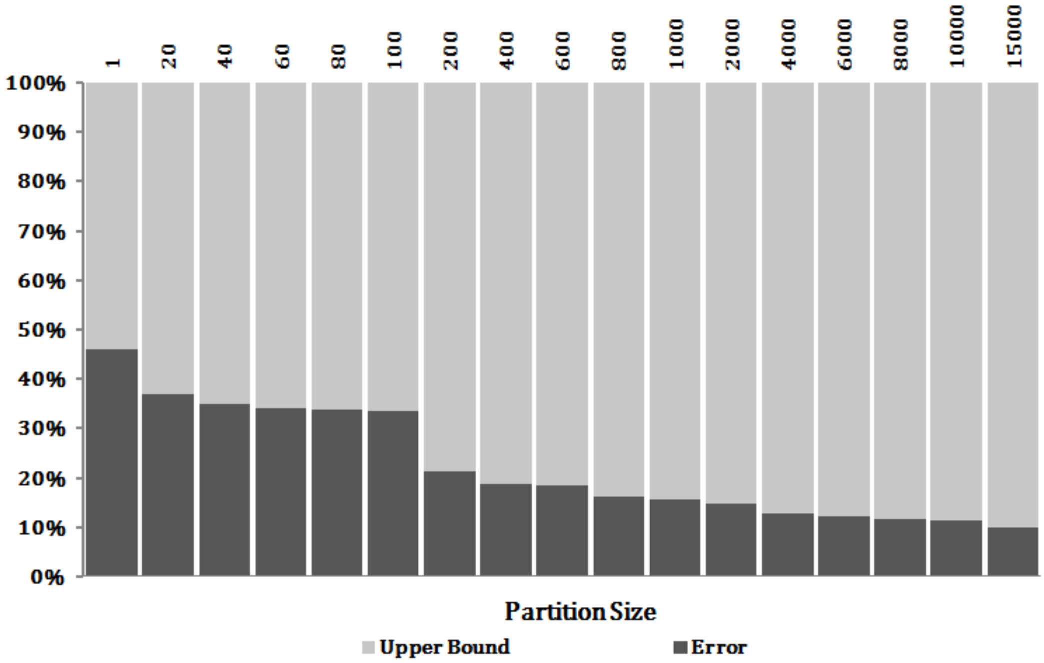



Figure 6.2: Upper and lower bounds estimation.



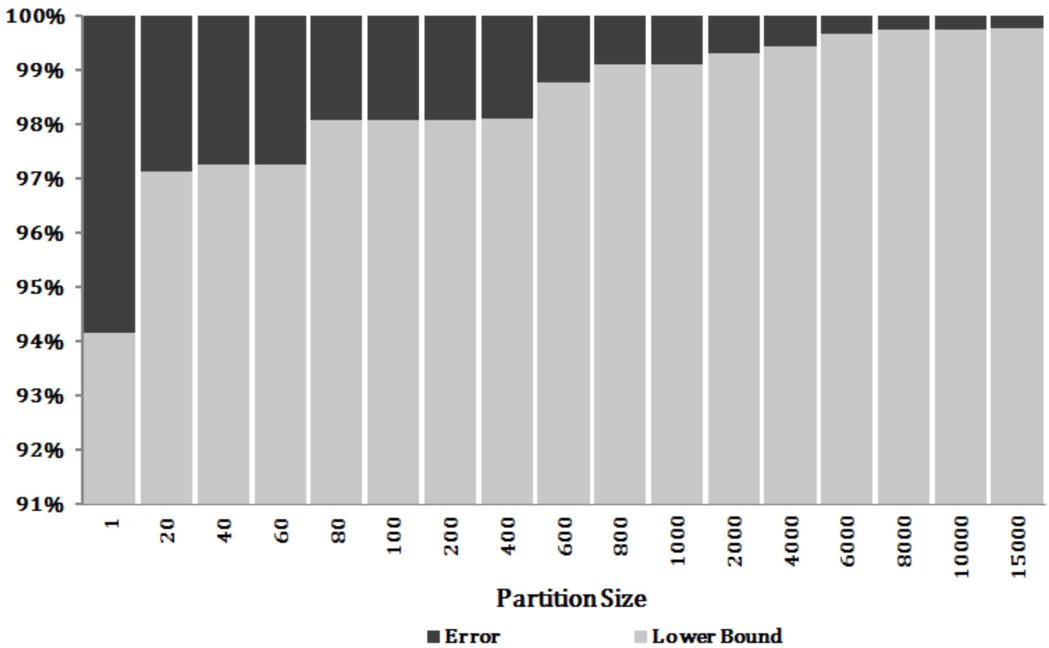




Figure 6.3: Error graph.

It is worth mentioning that the computational cost to estimate the upper bound, was higher than the computational cost to estimate the lower one, and this last was quite similar to the computational cost demanded by the SAA estimatior. This is reasonable since the linear programming problem (5-9) is equivalent to (2-26) considering the "mean scenarios" (centers of the cells) as sample and $\{P(\xi \in I_k)\}_{k=1}^{n}$ as discrete probability distribution.

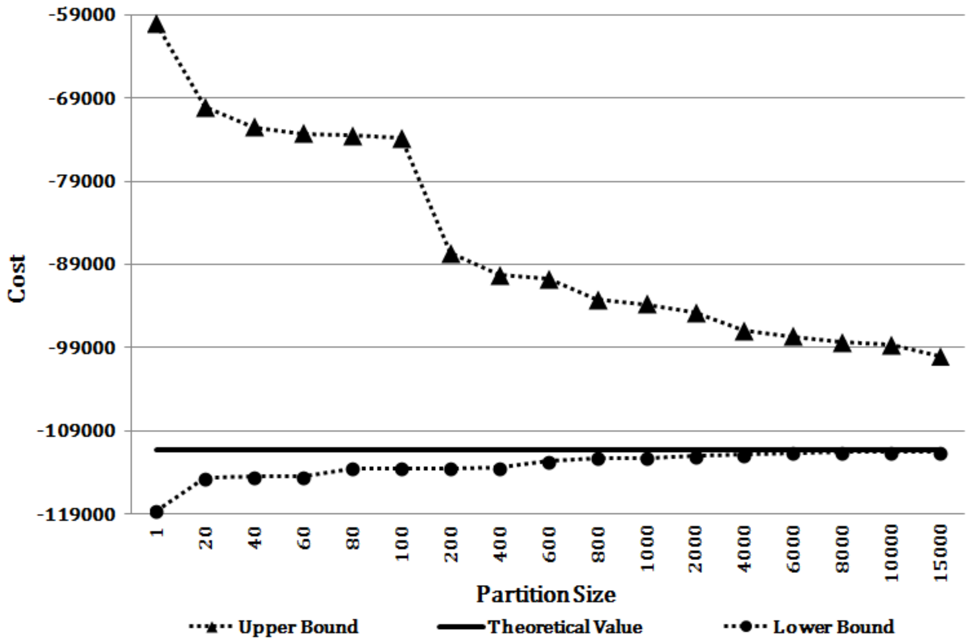



Figure 6.4: Comparison of upper and lower bound convergence.

Even though the computational cost of the partition-based method higher than SAA estimator one, this alternative technique is desirable for two-stage stochastic programming problems with convex and complete recourse function, since this method gives a deterministic estimation of the true solution and allows to get a more accurate approximation, increasing the information about the unknown data (increasing the partition size) to adjust the conservatism level of the model.

## 6.3 Numerical Results from the SAWPR

The obtained results from SAWPR algorithm are presented below:

| Partition Size[2] | Upper Bound | Lower Bound |
|---|---|---|
| 2 | -78400,00429 | -113554,5650 |
| 10 | -91000,00485 | -113065,7551 |
| 20 | -95250,02819 | -113065,6612 |
| 30 | -97256,25214 | -111753,9764 |
| 40 | -98100,00023 | -111753,9840 |
| 50 | -99295,31262 | -111641,4684 |
| 60 | -99663,88845 | -111632,3311 |
| 70 | -100060,2431 | -111632,3309 |
| 80 | -100429,1654 | -111632,3308 |
| 90 | -100834,9826 | -111632,3297 |
| 100 | -101019,4443 | -111632,2889 |
| 200 | -103062,4998 | -111516,1072 |
| 300 | -104317,4107 | -111396,4226 |
| 400 | -105094,8800 | -111396,4498 |
| 500 | -105636,1365 | -111396,4487 |
| 600 | -105896,3662 | -111396,4488 |
| 700 | -106151,8539 | -111396,4504 |
| 800 | -106417,3859 | -111396,4498 |
| 900 | -106542,9439 | -111396,4291 |

Table 6.2: Numerical results.

This data are graphicaly represented in the next figures where the convergence of both techniques (the SARPR and SAWPR algorithm) are compared.

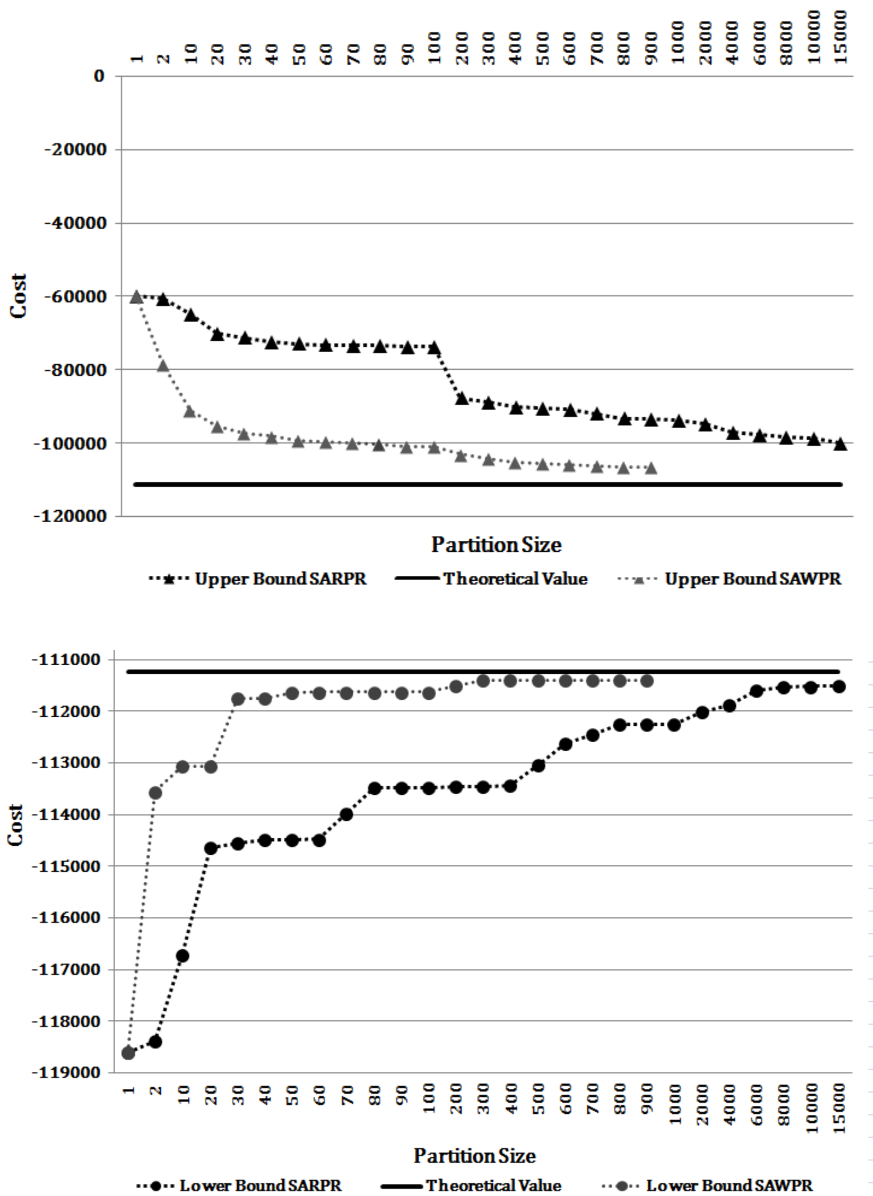



Figure 6.5: Bounding from SAWPR Algorithm.

Observing the figure above is clear that SAWPR algorithm converges for a significantly smaller partition size than SARPR algorithm.

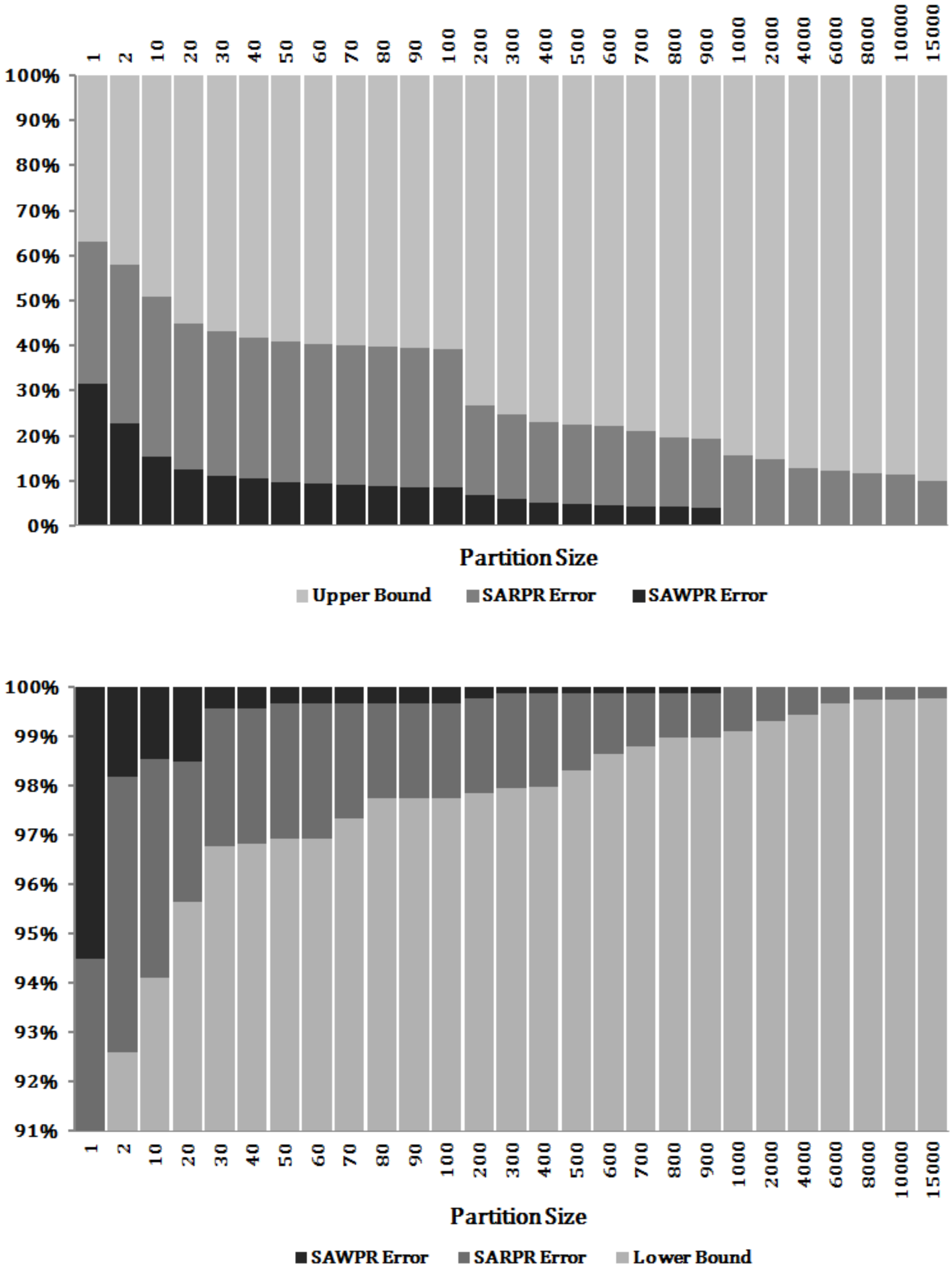



Figure 6.6: Estimation error.

The Figure 6.6 shows that the approximation error for SAWPR algorithm is less than the random method error for each partition size, which proves its numerical accuracy.

# 7
# Final remarks

- The SAA (Sample Average Approximation) is a useful method but has disadvantages, like: requires a big scenarios set to converge, which affects its computational efficiency; Its optimal solution depends on the sample, and that for that reason it is a random variable also.

- The partition-based method is convergent. Theoretically we can prove that if the recourse function is continuous and the uncertain variable has compact support, then the approximate solution converges to the optimal solution of the true problem.

- The performance-output of the partition-based method gives a deterministic estimation for the optimal solution of the true problem.

- The SAWPR algorithm shows that with robust optimization techniques it is possible to find a better way to partition the support of the uncertain variable to reach the optimal solution with a smaller partition size, and improve the computational efficiency.

- The computational results show that both bounds convergence at different velocities. The lower bound is less conservative than the upper one, but increases the information about the distribution of the uncertain variable (increasing the partition size), the robustness of the model adjust the conservatism level for the upper optimal solution.

# 8
# Future Works

– Improve the SAWPR algorithm by conssidering another way to choose the cutting point.

– Speed up the convergence of the refinement sequential problem, analyzing valid restrictions, i.e., identifying analytically the optimal solution for the worst-case problem.

– Consider the case $\xi := q$, when the recourse function is concave and the worst-case problem

$$\operatorname*{Max}_{\xi \in I_k} Q(x, \xi)$$

could be solved by the Benders' cut method or column-and-constraint generation method, for example.

– Test the partition-based method for more instances in order to identify the class of problems for which the method works.

– Compare the partition-based method with improved versions of the Sample Average Approximation method.

– Extend the partition-based method to multi-stage stochastic programming case.

# 9
# Bibliography


BERTSIMAS, D.; SIM, M. The price of robustness. **Operations Research**, Institute for Operations Research and the Management Sciences (INFORMS), Maryland, USA, vol. 52, no. 1, p. 35–53, Jannuary-February 2004. ISSN 0030-364X. Available from Internet: <http://dx.doi.org/10.1287/opre.1030.0065>. 4.3

BIRGE, J.; LOUVEAUX. **Introduction to Stochastic Programming**. Second. New York Dordrecht Heidelberg London: Springer, 2011. 2.6, 3.2

BIRGE, J. R.; WEST, R. J. Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. **Stochastic Programming 84 Part I**, p. 54–102, 1986. 1

CARØE, C. C.; TIND, J. L-shaped decomposition of two-stage stochastic programs with integer recourse. **Mathematical Programming**, vol. 83, no. 3, p. 451, 1998. 1

EDIRISINGHE, N.; ZIEMBA, W. T. Implementing bounds-based approximations in convex-concave two-stage stochastic programming. **Mathematical Programming**, vol. 75, no. 2, p. 295–325, 1996. 1

EDIRISINGHE, N. C. P.; ZIEMBA, W. T. Bounds for two-stage stochastic programs with fixed recourse. **Mathematics of Operations Research**, vol. 19, no. 2, p. 292–313, 1994.

ERMOLIEV, Y. M.; NORKIN, V. I. Sample average approximation method for compound stochastic optimization problems. **SIAM Journal on Optimization**, vol. 23, no. 4, p. 2231–2263, 2013.

GLYNN, P. W.; INFANGER, G. Simulation-based confidence bounds for two-stage stochastic programs. **Mathematical Programming**, vol. 138, no. 1-2, p. 15–42, 2013.

HUANG C. C., W. T. Z.; BEN-TAL, A. Bounds on the expectation of a convex function of a random variable: With applications to stochastic programming. **Operations Research**, vol. 25, no. 2, p. 315–325, 1977.

KLEYWEGT, A. J.; SHAPIRO, A.; MELLO, T. Homem-de. The sample average approximation method for stochastic discrete optimization. **SIAM Journal on Optimization**, vol. 12, no. 2, p. 479–502, 2002. 2.4

PHELPS CHRIS, J. O. R.; GONG, Q. Optimal control of uncertain systems using sample average approximations. **SIAM Journal on Optimization**, vol. 54, no. 1, p. 1–29, 2016.

PIERRE-LOUIS PÉGUY, G. B.; MORTONL, D. P. A combined deterministic and sampling-based sequential bounding method for stochastic programming. **Proceedings of the Winter Simulation Conference. Winter Simulation Conference**, p. 4172–4183, 2011.

ROCKAFELLAR, T. **Convex Analysis**. Princeton, New Yersey U.S.A.: Princeton University Press, 1970. 3.3.2

SHAPIRO, A.; DENTCHEVA, D.; RUSZCZYŃSK, A. **Lectures on stochastic programming**. Philadelphia: Society for industrial and applied mathematics, 2009. 2.1, 2.2.1, 2.5

SLYKE, R. M. V.; WETS, R. L-shaped linear programs with applications to optimal control and stochastic programming. **SIAM Journal on Applied Mathematics**, vol. 17, no. 4, p. 638–663, 1969.

SNIEDOVICH, M. **Dynamic Programming: foundations and principles**. Second. United States of America U.S.A: Taylor and Francis Group, LLC, 2011.

SONG, Y.; LUEDTKE, J. An adaptive partition-based approach for solving two-stage stochastic programs with fixed recourse. **SIAM Journal on Optimization**, vol. 25, no. 3, p. 1344–1367, 2015. 1

TRUKHANOV SVYATOSLAV, L. N.; SCHAEFER, A. Adaptive multicut aggregation for two-stage stochastic linear programs with recourse. **European Journal of Operational Research**, vol. 206, no. 2, p. 395–406, 2010.

TUY, H. **Convex Analysis and Convex Optimization**. Netherlands: Kluwer Academic Publishers, 1998. 3.3.2

# A
# Mathematical and Statistical Background

## A.1
## The Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with mean $\mu$ and finite variance $\sigma^2$. Then, for any $\epsilon > 0$, we have that

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2},$$

from which we obtain that for any $\epsilon > 0$:

$$\lim_{n \longrightarrow \infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) = 0.$$

***Proof.*** *Let $\overline{X}_n := \frac{X_1+\cdots+X_n}{n}$ be the arithmetic mean of the first $n$ random variables. Clearly $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{V}ar(\overline{X}_n) = \frac{\sigma^2}{n}$. Chebyschev's inequality implies that for any $\epsilon > 0$ we must have:*

$$P(|\overline{X}_n - \mathbb{E}[\overline{X}_n]| \geq \epsilon) \leq \frac{\mathbb{V}ar(\overline{X}_n)}{\epsilon^2},$$

*wich is exactly what we wanted to prove.* □

Note that

$$\lim_{n \longrightarrow \infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) = 0,$$

is equivalent to

$$\lim_{n \longrightarrow \infty} P\left(\frac{X_1 + \cdots + X_n}{n} = \mu\right) = 1.$$

## A.2
## The Central Limit Theorem

(**Univariate Case**) Let $X_1, X_2, \ldots$ be a sequence of independent and indentically distributed random variables with mean $\mu$ and finite variance $\sigma^2$. Let $S_n := \sum_{j=1}^{n} X_j$ and $Y_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then, the sequence of random variables $Y_1, Y_2, \ldots$ converges in distribution to a random variable $Y$ having a standard normal distribution. In others words:

$$\lim_{n\longrightarrow\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \quad \forall x \in \mathbb{R}.$$

***Proof*.** *Whithout loss of generality, it will be assumed that $\mu = 0$. Let $\phi$ be the characteristic function of the random variables $X_1, X_2, \ldots$ Since the random variables are independent and identically distributed, we have that:*

$$\begin{aligned}
\phi_{Y_n} &= \mathbb{E}[exp(itY_n)] \\
&= \mathbb{E}\left[exp\left(it\frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}}\right)\right] \\
&= \mathbb{E}\left[\prod_{j=1}^{n} exp\left(it\frac{X_j}{\sigma\sqrt{n}}\right)\right] \\
&= \prod_{j=1}^{n} \phi\left(\frac{t}{\sigma\sqrt{n}}\right) \\
&= \left[\phi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n
\end{aligned}$$

*Expansion of $\phi$ in a Taylor series about zero yields:*

$$\begin{aligned}
\phi(t) &= \mathbb{E}[e^{itX}] \\
&= \mathbb{E}\left[1 + itX - \frac{t^2X^2}{2} + \frac{it^3X^3}{6} - \cdots\right] \\
&+ 1 + 0 - \frac{\sigma^2 t^2}{2} + t^2 O_p(t).
\end{aligned}$$

*Thus:*

$$\begin{aligned}
\phi_{Y_n}(t) = &\left[1 - \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \left(\frac{t}{\sigma\sqrt{n}}\right)^2 O_p(t)\right]^n \\
&= exp\left[nln\left(1 - \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \left(\frac{t}{\sigma\sqrt{n}}\right)^2 O_p(t)\right)\right].
\end{aligned}$$

*Taking the limit when $n \to \infty$ we get:*

$$\lim_{n\longrightarrow\infty} \phi_{Y_n}(t) = exp\left[-\frac{t^2}{2}\right].$$

*The Lévy-Cramer continuity theorem implies that*

$$Y_n \xrightarrow[n\to\infty]{d} Y,$$

*where $Y$ is a random variable having a standard normal distribution.* $\square$

(**Multivariate Case**) The central limit theorem can be generalized to sequences of random vectors as follows:

Let $X_1, X_2, \ldots$ be sequence of $d$-dimensional independent and identically distributed random vectors ($X_j \in \mathbb{R}^d \quad \forall j$), having mean vector $\mu$ and variance-covariance matrix $\Sigma$, where $\Sigma$ is positive definite. Let

$$\overline{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

be the vector of arithmetic means. Then

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow[n\to\infty]{d} X,$$

where $X$ is a $d$-dimensional random vector having a multivariate normal distribution with mean 0 and variance-covariance matrix $\Sigma$.

## A.3 Continuity of Convex Functions



Let $U \subset \mathbb{R}^n$ be a open and convex set. Every convex function $f : U \longrightarrow \mathbb{R}$ is continuous.

The proof of this fact is supported on the next two propositions

**Proposition A.3.1.** *Every point of the rectangular bloc $B = \prod_{i=1}^{n}[a_i, b_i]$, is a convex combination of its vertexs.*

***Proof*.** *(By induction) Clearly the result holds for $n = 1$. Let $n > 1$. The block's vertexs are the $2^n$ elements of the set $\prod_{i=1}^{n}\{a_i, b_i\}$. An arbitrary point of the block can be write as $p = (x, y)$, where $y \in [a_n, b_n]$ and $x$ belongs to the block $B' = \prod_{i=1}^{n-1}[a_i, b_i]$, of $n-1$ dimension. By induction hypothesis, $x = \sum \alpha_j u_j$ is a convex combination of the vertexs $u_j \in B'$. The vextex of $B$ are of the form $v_j = (u_j, a_j)$ and $\bar{v}_j = (u_j, b_n)$. Putting $p_0 = (x, a_n)$ and $p_1 = (x, b_n)$, we have that $p_0 = \sum \alpha_j v_j$ and $p_1 = \sum \alpha_j \bar{v}_j$. Moreover, $y = (1-t)a_n + tb_n$, with $t = \frac{y-a_n}{b_n-a_n}$, so $p = (1-t)p_0 + tp_1 = \sum(1-t)\alpha_j v_j + \sum t\alpha_j \bar{v}_j$, which express the arbitrary point $p$ of the block $B$ as a convex combination of its vertexs.* $\square$

**Proposition A.3.2.** *Every convex function $f : U \longrightarrow \mathbb{R}$, difined on a open and convex set $U \subset \mathbb{R}^n$, is locally increased by a constant.*

***Proof*.** *Let $A = \prod_{i=1}^{n}(a_i, b_i)$ the interior of the rectangular block contents in $U$. If $w_j, \quad j = 1, \ldots, 2^n$, denote the vertexs of $A$ we have, for any $x \in A$, that $x = \sum \alpha_j w_j$ then, by the convexity of $f$, $f(x) \leq \sum \alpha_j \cdot f(w_j) \leq M$, where $M = \max_j\{f(w_j)\}$.* $\square$

***Proof.*** ***(Continuity of Convex Functions)*** *To simplify the notation, to proof the continuity of the function $f$ at arbitrary point $a \in U$, we can assume that $a = 0$ and $f(0) = 0$, since the set $U_0 = \{x \in \mathbb{R}^n | a - x \in U\}$ is convex, open, contents $0$ and the function $g : U_0 \longrightarrow \mathbb{R}$, difined by $g(x) = f(a - x) - f(a)$, satifies $g(0) = 0$, is convex and continuous at point $0$ if, and only if, $f$ is continuous at point $a$. By* ***Proposition A.3.2*** *, there is $c > 0$ and $M > 0$ such that $|x| \leq c \Longrightarrow f(x) \leq M$. Let $\epsilon > 0$ be given, without loss of generality we can assume that $\epsilon < M$. By the convexity of $f$ we have that*

$$f\left(\frac{\epsilon}{M}x\right) = f\left(\left(1 - \frac{\epsilon}{M}\right) \cdot 0 + \frac{\epsilon}{M}x\right) \leq \frac{\epsilon}{M} \cdot f(x),$$

*so*

$$f(x) \leq \frac{\epsilon}{M} f\left(\frac{M}{\epsilon}x\right).$$

*Taking $\delta = \frac{\epsilon c}{M}$, we get*

$$|x| < \frac{\epsilon c}{M} \Longrightarrow \left|\frac{M}{\epsilon}x\right| < c \Longrightarrow f\left(\frac{M}{\epsilon}x\right) \leq M \Longrightarrow f(x) \leq \epsilon.$$

*Moreover,*

$$\begin{aligned} 0 = f(0) &= f\left(\frac{M}{M+\epsilon}x + \frac{\epsilon}{M+\epsilon}\left(-\frac{M}{\epsilon}x\right)\right) \\ &\leq \frac{M}{M+\epsilon} f(x) + \frac{\epsilon}{M+\epsilon} f\left(-\frac{M}{\epsilon}x\right). \end{aligned}$$

*Simplifying, we have that $M \cdot f(x) + \epsilon \cdot f(-Mx/\epsilon) \geq 0$, so:*

$$f(x) \geq \frac{\epsilon}{M} \cdot (-f(-Mx/\epsilon)) \geq \frac{\epsilon}{M} \cdot (-M) = -\epsilon.$$

*Thus, $|x| < c\epsilon/M \Longrightarrow -\epsilon \leq f(x) \leq \epsilon$, and therefore $f$ is continuous at point* $0$. $\square$

## A.4
## Jensen's Inequality

Let $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ be a convex function. If $X$ is a random variable and integrable, i.e. $\mathbb{E}[X] < \infty$, then

$$\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X]).$$

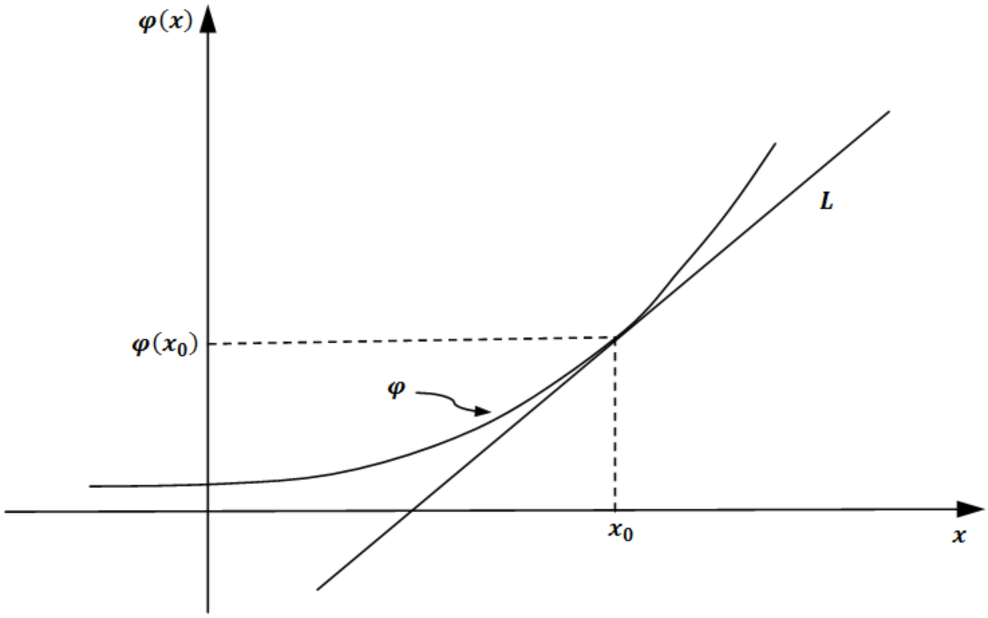



Figure A.1: Convex function.

***Proof.*** *Given $x_0$ and the point $\varphi(x_0)$ in the graph $G = \{(x, \varphi(x))|x \in \mathbb{R}\} \subset \mathbb{R}^2$, by convexity, there is a line $L$ that crosses by the point $(x_0, \varphi(x_0))$ and satisfies $\varphi(x) \geq L(x)$, $\forall x \in \mathbb{R}$.*

*Let $y - \varphi(x_0) = \lambda(x - x_0)$ the equation for the line $L$ for some $\lambda \in \mathbb{R}$. So*

$$\varphi(x) \geq L(x) = \varphi(x_0) + \lambda(x - x_0), \quad \forall x.$$

*Therefore,*

$$\mathbb{E}[\varphi(X)] \geq \mathbb{E}[L(X)] = \varphi(x_0) + \lambda(\mathbb{E}[X] - x_0).$$

*Putting $x_0 = \mathbb{E}[X]$, we get $\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X])$.* $\square$

### A.4.1 Jensen's Inequality for Conditional Expectation

Let $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ be a convex function. Then for any $X$, such that $\mathbb{E}[\varphi(X)] < \infty$,

$$\mathbb{E}[\varphi(X)|\mathscr{G}] \geq \varphi(\mathbb{E}[X|\mathscr{G}]) \quad \text{almost surely (a.s.)}$$

***Proof*.** *Define* $\mathscr{L} := \{(a,b) \in \mathbb{R}^2 | \, \varphi(x) > ax + b, \quad \forall x \in \mathbb{R}\}$. *By convexity* $\mathscr{L} \neq \emptyset$, *and also is open in the topology of* $\mathbb{R}^2$. *In particular,* $\mathscr{L} \cap \mathbb{Q}^2$ *is dense in* $\mathscr{L}$. *Then*

$$\varphi(x) = \underset{(a,b)\in\mathscr{L}\cap\mathbb{Q}^2}{sup} (ax+b), \quad x \in \mathbb{R}.$$

*Since* $\varphi(X) \geq aX + b$, *for each* $(a,b) \in \mathscr{L}$, *and since* $\varphi(X)$ *and* $X$ *are integrables, we have that*

$$\mathbb{E}[\varphi(X)|\mathscr{G}] \geq a\mathbb{E}[X|\mathscr{G}] + b \quad a.s.$$

*If* $\Omega_{a,b}$ *denote the event where the inequality holds, then* $P(\Omega_{a,b}) = 1$ *for all* $(a,b) \in \mathscr{L}$. *Define* $\Omega' := \bigcap_{(a,b)\in\mathscr{L}\cap\mathbb{Q}^2} \Omega_{a,b}$, *and note that* $P(\Omega') = 1$. *Taking the supremum over* $(a,b) \in \mathscr{L} \cap \mathbb{Q}^2$, *we have that*

$$\mathbb{E}[\varphi(X)|\mathscr{G}] \geq \varphi(\mathbb{E}[X|\mathscr{G}]), \quad on\, \Omega'.$$

□



## A.5 The Total Probability Law

Let $\{I_k\}_{k=1}^n$ be a partition of $\xi$ distribution function support. Since $\xi \in \Xi$ and $\bigcup_{k=1}^n I_k = \Xi$, for any $x$, it holds the following[1]:

$$\begin{aligned}
\mathbb{E}[Q(x,\xi)] =& \mathbb{E}\left[Q(x,\xi)\overbrace{\chi_{\bigcup\limits_{k=1}^{n} I_k}(\xi)}^{=1}\right] \\
=& \mathbb{E}\left[Q(x,\xi)\left(\sum_{k=1}^{n}\chi_{I_k}(\xi) - \sum_{k>1}\chi_{\bigcap\limits_{1\leq i_1<\cdots<i_k\leq n} I_{i_k}}(\xi)\right)\right].
\end{aligned}$$

[1] $\chi_A(x)$, denote the characteristic function, defined by

$$\chi_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

This last since

$$\chi_{\bigcup\limits_{k=1}^{n} I_k}(\xi) = \sum_{k=1}^{n} \chi_{I_k}(\xi) - \sum_{k>1} \chi_{\bigcap\limits_{1\le i_1<\cdots<i_k\le n} I_{i_k}}(\xi),$$

so

$$\begin{aligned}
&\mathbb{E}\left[Q(x,\xi)\left(\sum_{k=1}^{n} \chi_{I_k}(\xi) - \sum_{k>1} \chi_{\bigcap\limits_{1\le i_1<\cdots<i_k\le n} I_{i_k}}(\xi)\right)\right] \\
&= \mathbb{E}\left[Q(x,\xi)\sum_{k=1}^{n} \chi_{I_k}(\xi)\right] - \overbrace{\mathbb{E}\left[Q(x,\xi)\sum_{k>1} \chi_{\bigcap\limits_{1\le i_1<\cdots<i_k\le n} I_{i_k}}(\xi)\right]}^{=0} \\
&= \mathbb{E}\left[\sum_{k=1}^{n} Q(x,\xi)\chi_{I_k}(\xi)\right] \\
&= \sum_{k=1}^{n} \mathbb{E}[Q(x,\xi)\chi_{I_k}(\xi)] = \sum_{k=1}^{n} \frac{\mathbb{E}[Q(x,\xi)\chi_{I_k}(\xi)]}{P(\xi\in I_k)} P(\xi\in I_k),
\end{aligned}$$

[2]and by deffinition,

$$\frac{\mathbb{E}[Q(x,\xi)\chi_{I_k}(\xi)]}{P(\xi\in I_k)} = \mathbb{E}[Q(x,\xi)|\xi\in I_k],$$

thus,

$$\mathbb{E}[Q(x,\xi)] = \sum_{k=1}^{n} \mathbb{E}[Q(x,\xi)|\xi\in I_k]P(\xi\in I_k). \tag{A-1}$$

## A.6 Large Linear Programming Problem for the Upper Bound

If $(\theta_k^*, y^*, w^*)$ is the optimal solution for (5-4), and if we suposse that there is $s' \in \{-,+\}^3$, such that $Q(x,\xi^{(k,s')}) > \theta_k^*$, then

$$\begin{cases}
\xi_T^{(k,s')} x_T + y_T^{*(k,s')} - w_T^{*(k,s')} \ge 200, \\
\xi_M^{(k,s')} x_M + y_M^{*(k,s')} - w_M^{*(k,s')} \ge 240, \\
\xi_B^{(k,s')} x_B - w_{B_1}^{*(k,s')} - w_{B_2}^{*(k,s')} \ge 0, \\
- w_{B_1}^{*(k,s')} \ge -6000, \\
y_T^{*(k,s')}, y_M^{*(k,s')} \ge 0, \quad w_T^{*(k,s')}, w_M^{*(k,s')}, w_{B_1}^{*(k,s')}, w_{B_2}^{*(k,s')} \ge 0,
\end{cases}$$

and

[2] $\mathbb{E}\left[Q(x,\xi)\sum_{k>1}\chi_{\bigcap\limits_{1\le i_1<\cdots<i_k\le n} I_{i_k}}(\xi)\right] = 0$, since $P\left(\xi\in \bigcap\limits_{1\le i_1<\cdots<i_k\le n} I_{i_k}\right) = 0$ for all $k>1$

$$Q(x,\xi^{(k,s')}) > 238y_T^{*(k,s')}+210y_M^{*(k,s')}-170w_T^{*(k,s')}-150w_M^{*(k,s')}-36w_{B_1}^{*(k,s')}-10w_{B_2}^{*(k,s')},$$

contradicting the minimality of $Q(x,\xi^{(k,s')})$.

By other hand, if we suposse that there is $s' \in \{-,+\}^3$, such that

$$\begin{aligned}\theta_k^* &= 238y_T^{*(k,s')} + 210y_M^{*(k,s')} - 170w_T^{*(k,s')} - 150w_M^{*(k,s')} - 36w_{B_1}^{*(k,s')} - 10w_{B_2}^{*(k,s')}\\ &> Q(x,\xi^{(k,s')}),\end{aligned}$$

then by existence and optimality of $Q(x,\xi^{(k,s')})$, there is $(\tilde{\theta}_k\tilde{y},\tilde{w})$, such that

$$\begin{cases}\xi_T^{(k,s')}x_T + \tilde{y}_T^{(k,s')} - \tilde{w}_T^{(k,s')} \geq 200,\\ \xi_M^{(k,s')}x_M + \tilde{y}_M^{(k,s')} - \tilde{w}_M^{(k,s')} \geq 240,\\ \xi_B^{(k,s')}x_B - \tilde{w}_{B_1}^{(k,s')} - \tilde{w}_{B_2}^{(k,s')} \geq 0,\\ -\tilde{w}_{B_1}^{(k,s')} \geq -6000,\\ \tilde{y}_T^{(k,s')}, \tilde{y}_M^{(k,s')} \geq 0, \quad \tilde{w}_T^{(k,s')}, \tilde{w}_M^{(k,s')}, \tilde{w}_{B_1}^{(k,s')}, \tilde{w}_{B_2}^{(k,s')} \geq 0,\end{cases}$$

and

$$\begin{aligned}\tilde{\theta}_k &= 238\tilde{y}_T^{(k,s')} + 210\tilde{y}_M^{(k,s')} - 170\tilde{w}_T^{(k,s')} - 150\tilde{w}_M^{(k,s')} - 36\tilde{w}_{B_1}^{(k,s')} - 10\tilde{w}_{B_2}^{(k,s')}\\ &= Q(x,\xi^{(k,s')}).\end{aligned}$$

Therefore $\theta_k^* > \tilde{\theta}_k$, contradicting the optimality of $\theta_k^*$.



## A.7 Average Optimizer

Put

$$\begin{aligned}\theta_k^* = \text{Min} \quad & \theta_k\\ \text{s.t.} \quad & \theta_k \in \Theta_k,\end{aligned}$$

where $\Theta_k$ denote the set of restrictions for the problem (5-4), and

$$\begin{aligned}\tilde{\theta} = \text{Argmin} \quad & \sum_{k=1}^{n} \theta_k \cdot P(\xi \in I_k)\\ \text{s.t.} \quad & \theta_k \in \Theta_k, \quad \forall k,\end{aligned}$$

then $\tilde{\theta}_k \in \Theta_k$, thus

$$\Longrightarrow \tilde{\theta}_k \geq \theta_k^*$$
$$\Longrightarrow \sum_{k=1}^{n} \tilde{\theta}_k \cdot P(\xi \in I_k) \geq \sum_{k=1}^{n} \theta_k^* \cdot P(\xi \in I_k).$$

If we suposse that there is $k'$ such that $\tilde{\theta}_{k'} > \theta_{k'}^*$, then we can replace $\tilde{\theta}_{k'}$ by $\theta_{k'}^*$ and get the vector $(\tilde{\theta}_1, \ldots, \tilde{\theta}_{k'-1}, \theta_{k'}^*, \tilde{\theta}_{k'+1}, \ldots, \tilde{\theta}_n)$, so

$$\sum_{k=1}^{n} \tilde{\theta}_k \cdot P(\xi \in I_k) > \sum_{k=1}^{k'-1} \tilde{\theta}_k \cdot P(\xi \in I_k) + \theta_{k'}^* \cdot P(\xi \in I_{k'}) + \sum_{k=k'+1}^{n} \tilde{\theta}_k \cdot P(\xi \in I_k),$$

contradicting the optimality of $\tilde{\theta} = (\tilde{\theta}_k)_{k=1}^n$.

## A.8 Edmundson-Madansky Inequality

Let $ext\Xi$ be the set of extreme points of the convex hull of $\Xi$, and $\mathcal{E}$ the sigma algebra of all subsets of $ext\Xi$.

Suposse that $\xi \longmapsto Q(x,\xi)$ is convex and $\Xi$ is compact. For all $\xi \in \Xi$ let $\phi(\xi, \cdot)$ be a probability measure on $(ext\Xi, \mathcal{E})$, such that

$$\int_{e \in ext\Xi} e\phi(\xi, de) = \xi, \tag{A-2}$$

and $\omega \longmapsto \phi(\xi(\omega), A)$ is measurable for all $A \in \mathcal{E}$. Then

$$\mathbb{E}[Q(x,\xi)] \leq \int_{e \in ext\Xi} Q(x,e)\lambda(de), \tag{A-3}$$

where $\lambda$ is the probability measure on $\mathcal{E}$ defined by

$$\lambda(A) = \int_{\Omega} \phi(\xi(\omega), A)P(d\omega). \tag{A-4}$$

***Proof.*** *Since $\omega \longmapsto Q(x, \xi(\omega))$ is convex, and*

$$\xi = \int_{e \in ext\Xi} e \cdot \phi(\xi, de) \quad \forall \xi \in \Xi,$$

*is a convex combination of extreme points, it holds*

$$\begin{aligned} Q(x,\xi) = &\ Q(x, \int_{e \in ext\Xi} e \cdot \phi(\xi, de)) \\ \leq &\int_{e \in ext\Xi} Q(x,e) \cdot \phi(\xi, de), \end{aligned}$$

*so integrating with respect to probability measure $P$, we have*

$$\begin{aligned}
\mathbb{E}[Q(x,\xi)] &= \int_{\omega\in\Omega} Q(x,\xi(\omega))P(d\omega) \\
&\leq \int_{\omega\in\Omega}\int_{e\in ext\Xi} Q(x,e)\cdot\phi(\xi(\omega),de)P(d\omega) \\
&= \int_{e\in ext\Xi} Q(x,e)\underbrace{\int_{\omega\in\Omega}\phi(\xi(\omega),de)P(d\omega)}_{\lambda(de)} \\
&= \int_{e\in ext\Xi} Q(x,e)\lambda(de).
\end{aligned}$$

□

The compatibility relation between the probability measure $\lambda$ and the distribution of $\xi$ (see 3.2), allows to indentify the set where we can find the optimal measure $\lambda$ that satisfies the conditions of the Edmundson-Madansky inequality, since

$$\lambda \in \{\mu \mid \mu \,\text{is a probability measure on}\, \mathcal{E}, \text{and}\, \mathbb{E}_{\mu}[e] = \bar{\xi}\}.$$