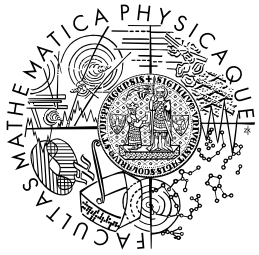

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

**MASTER THESIS**

Erik Kočandrle

# Stochastic programming problems with endogenous randomness

Department of Probability and Mathematical Statistics

Supervisor of the master thesis: doc. RNDr. Ing. Miloš Kopa, Ph.D.
Study programme: Probability, Mathematical Statistics and Econometrics

Prague 2024

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date .....  .....

Author's signature

I would like to dedicate this thesis to my supervisor doc. RNDr. Ing. Miloš Kopa, Ph.D. and thank him for patiently mentoring me through this qualification paper. It has been an honor to go through this journey alongside such a skilled and experienced mathematician. I also need to thank all who understood and supported me through these challenging months, my family, friends and girlfriend.

Title: Stochastic programming problems with endogenous randomness

Author: Erik Kočandrle

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Ing. Miloš Kopa, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This thesis builds upon the standard stochastic programming model and introduces an extension for endogenous randomness, which can be more suitable in certain situations. The classical model of stochastic programming assumes that the distribution of the underlying random element is independent of the decision makers actions. This assumption is not always reasonable and practical problems often include situations, where the decisions maker can alter the distribution in various ways. Models with endogenous randomness often exhibit high complexity and computational expense. This thesis summarizes several tractable modelling techniques with motivational examples demonstrating their possible use. The derived principles are then demonstrated by providing extensions of the classical newsvendor problem and the CVaR portfolio optimization problem, where we assume that a large enough investment can alter the assets loss distribution.

Keywords: stochastic programming, multi-stage programming, endogenous randomness, newsvendor problem, portfolio optimization, CVaR risk measure

Název práce: Úlohy stochastického programování s endogenní náhodou

Autor: Erik Kočandrle

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí práce: doc. RNDr. Ing. Miloš Kopa, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Tato práce staví na standardním modelu stochastického programování a představuje rozšíření pro endogenní náhodu, které může být v určitých situacích vhodnější. Klasická úloha stochastického programování předpokládá, že rozdělení náhodného elementu nezávisí na rozhodnutích. Tento předpoklad není vždy splněný a v praktických úlohách často potkáváme situace, kde subjekt může různými způsoby toto rozdělení ovlivnit. Modely s endogenní náhodou často vykazují vysokou míru komplexity a výpočetní složitosti. Tato práce shrnuje několik užitečných technik pro modelování endogenní náhody spolu s příklady demonstrujícími jejich možná využití. Zkoumané principy jsou poté demonstrovány rozšířením klasického problému prodavače novin a úlohy optimalizace portfolia pomocí rizikové míry CVaR, kde předpokládáme, že dostatečně velká investice dokáže pozměnit rozdělení ztrát jednotlivých aktiv.

Klíčová slova: stochastické programování, vícestupňové programování, endogenní náhoda, problém prodavače novin, optimalizace portfolia, CVaR míra rizika

# Contents

# Introduction

Mathematical programming is a complex and versatile field with a wide array of possible applications. The standard programming model assumes, that all the quantities entering the model are known by the decision maker. This is often not the case, since the real world is full of uncertainty. Simply replacing the random element by its estimated value might result in a loss of important information about the system at hand. A more refined approach is to utilize the stochastic programming framework working directly with the probability distribution, which is either known or estimated from available data. There is usually an implicit assumption included with these models, that the underlying distribution is fixed and does not depend on the decisions made within the model. There are many cases, where it is not a valid assumption, usually when the uncertainty comes from within the system. This phenomenon is called endogenous randomness (or decision dependent randomness), as opposed to the classical case of exogenous randomness. The actions of the decision maker can therefore alter the underlying probability distribution in many different ways. Not only can the decisions influence the distribution itself, but in some multi-stage models they can also change the times at which the decision maker obtains new information about the random element.

Some mentions of endogenous randomness in stochastic programming could be traced more than 70 years back, however, it is still a relatively new topic in literature. One of the earlier works can be found in Pflug (1990), where endogenous randomness was incorporated into optimization of Markovian processes. One of the first more general works is Jonsbråten et al. (1998), where a new class of problems was created. Through the years, mathematicians were searching for special problem structures, which allow for efficient modelling of endogenous randomness. A large multi-stage model was presented in Goel and Grossmann (2004), which created an optimization framework for a large offshore oil field. The paper Dupačová (2006) mentioned the use of contamination and its importance for stress testing. Lot of works were interested in robust formulations of these problems, for example Basciftci et al. (2021) or Luo and Mehrotra (2020), which studied the concept of ambiguity sets. Another robust technique was contamination, presented for example in Kopa and Rusý (2023). Searching for tractable problem structures is still an ongoing process in literature. In the paper we also present the newsvendor problem, which can be found in Hrabec et al. (2012) for example, and the important $CVaR$ risk measure found in Pflug (2000).

There are infinitely many options for how a decision variable can transform a probability distribution. The goal of this thesis is not to provide an exhaustive list of them, but rather to try and find various tractable cases, which are computationally viable and make sense in relation to possible applications.

The theoretical part of the thesis consists of the first two chapters. The first chapter contains a summary of important stochastic optimization aspects, including the basics of multi-stage problems. The second chapter is trying to set

up a baseline mathematical foundation for endogenous randomness modelling. The important concepts from classical stochastic programming with exogenous randomness are extended and new types of problems are formulated for different types of decision dependence. There are simple motivational examples included, demonstrating the possible situations where such models can occur. The main goal of the second chapter is to find and summarize important cases of endogenous randomness using a unified notation and mathematical background.

In chapters 3 and 4 we extended the standard newsvendor model to include decision dependent demand and advertisement. After that we came up with a three-stage formulation with recourse decisions. Both models are then presented on a simple numerical example. Chapter 5 is dedicated to portfolio optimization with respect to the CVaR risk measure. We included decision dependent randomness in the asset losses after large investments into a company are made. The goal of these models is merely to demonstrate the important concepts from chapter 2. The values entering them are not intended to be perfectly realistic, however we tried choosing them in a reasonable way, which does not violate fundamental concepts of economics and finance.

# 1. Stochastic programming

## 1.1 Short introduction

The goal of the field of mathematical programming is to search for optimal (or at least acceptable) decisions in various problems. The classical case consists of an objective function $f : \mathbb{R}^n \to \mathbb{R}$ and a set of feasible solutions $\mathcal{X}_0$. In this thesis we will implicitly work with minimization problems, unless stated otherwise. The maximization problem can always be easily transformed to a minimization problem. The general formulation of a classical deterministic programming problem is

$$\min_{x \in \mathcal{X}_0} f(x).$$

Throughout the thesis, we will consider the feasibility set $\mathcal{X}_0$ to be closed, unless stated otherwise. The deterministic programming framework comes in several possible variations and is capable of handling many real world problems. However, it assumes that all parameters within the model are known in advance (i.e. before any decisions are made). In a world full of uncertainty, there are situations where this assumption is not sound and another approach must be taken. This is where the stochastic programming framework comes in.

## 1.2 Stochastic programming model

As mentioned earlier, there are cases where we encounter uncertain parameters while constructing the optimization model. The most basic approach would be to simply replace the random element by some sort of estimate and solve the problem deterministically. However, that approach has many limitations. We will start building the mathematical representation by defining a probability space $(\Omega, \mathcal{A}, \mathcal{P}_\Omega)$. Let the uncertainty in the problem be represented by a d-dimensional real random vector $\boldsymbol{\xi} : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}^d)$ ($\mathcal{B}$ being the Borel $\sigma$-algebra) with a probability distribution $\mathcal{P}$ and a support $\Xi \subseteq \mathbb{R}^d$, meaning the smallest set such that $\mathbb{P}(\boldsymbol{\xi} \in \Xi) = 1$. A realization of the random vector $\boldsymbol{\xi}$ will be denoted as $\xi$, this logic will be used throughout the whole thesis. Note that the random element in the model does not necessarily have to be a numerical vector. For example, the outcomes of a coin toss are either heads or tails, but we can code them numerically as 0 and 1. Generally we will work with real vectors, but even when it is not the case, we assume that we can code the scenarios by using numerical values.

Stochastic programming generally handles situations, where the objective function or the constraints are somehow dependent on the realization of the random

element. This makes the problem more nuanced apart from the deterministic one. In the end we always have to reformulate the problems deterministically, but there are several different ways to do that. The standard framework of stochastic programming includes one core assumption, which is rarely explicitly stated. However, it needs to be stated now, since the primary topic of this thesis is the case where the assumption does not hold. The core assumption is, that the distribution $\mathcal{P}$ of the vector $\boldsymbol{\xi}$ is fixed, in other words, it does not depend on the decisions. This way we can think of the random element as an external influence, which can not be affected by the decision maker and the decisions need to be adjusted to it. This is the case of **exogenous randomness** and as we will see in the next chapter, it is not always a good assumption to make. Throughout this chapter, we will consider that it holds.

### 1.2.1 Model formulation

The general uncertain form of the problem is

$$\text{"}\min_{x\in\mathcal{X}(\boldsymbol{\xi})} f(x,\boldsymbol{\xi})\text{"}, \tag{1.1}$$

where the objective function and feasibility set depend on the random element. Such uncertain formulations will be notated using quotation marks. For the sake of simplicity, we will mostly work with the case (1.2), where the randomness in the feasibility set is present through a finite number of equalities and inequalities. This will be the main definition of a stochastic program in this thesis, generally we would call it a nonlinear stochastic program.

**Definition 1.** *Using the notations above, we define a* ***Stochastic programming model*** *by a following uncertain formulation:*

$$\begin{gathered}\text{"}\min_{x\in\mathcal{X}_0}\ f(x,\boldsymbol{\xi})\text{"}\\ s.t.\\ \text{"}g_j(x,\boldsymbol{\xi})\le 0\text{"},\ j=1,...,p,\\ \text{"}h_k(x,\boldsymbol{\xi})=0\text{"},\ k=1,...,q,\end{gathered} \tag{1.2}$$

*where $p,q\in\mathbb{N}_0$, $f,g_j,h_k:\mathbb{R}^n\times\Xi\to\mathbb{R}$ are given functions $\forall i,j$ and $\mathcal{X}_0\subseteq\mathbb{R}^n$ is a set of hard constraints, which do not depend on the random element.*

*Remark.* In relation to the general formulation (1.1), for each realization $\xi\in\Xi$ is the feasibility set defined as

$$\mathcal{X}(\xi)=\{x\in\mathcal{X}_0 : g_j(x,\xi)\le 0\ \forall j,\ h_k(x,\xi)=0\ \forall k\}.$$

The uncertain formulation with random elements needs to be reformulated deterministically, for which there are various methods. For simplicity, we will distinguish two cases, when the randomness is only in the objective function or only in the constraints. If it is present in both, we just combine these methods.

### 1.2.2 Randomness in the objective function

We will now assume that the set of feasible decisions does not depend on the random parameter. Our goal is to reformulate the uncertain problem

$$\text{"}\min_{x\in\mathcal{X}_0} f(x,\boldsymbol{\xi})\text{"}.$$

We can look at this in the scope of multi-objective optimization. Basically, the decision maker would like to minimize all the functions $f(x,\xi)$ for all $\xi \in \Xi$, with respect to x. The perfect solution would minimize the objective for all possible realizations, but that is rarely possible. The simple way to handle this issue would be to turn it into a deterministic programming model by replacing the random parameter $\boldsymbol{\xi}$ by some estimate $\hat{\xi}$ from available data. This would result in the problem $\min_{x\in\mathcal{X}_0} f(x,\hat{\xi})$. This is not an ideal approach, since we lose valuable information about the underlying random element. For example, if we replace the parameter by its expected value, the result does not even have to lie inside the support $\Xi$. Unfortunately, this approach may be necessary when the stochastic makes the problem too computationally expensive.

Another approach of reformulating the problem is to use **robust optimization**. This can be plausible when the decision maker wants to hedge against the worst possible scenario. The method of choice is the min-max formulation shown in the following program

$$\min_{x\in\mathcal{X}_0} \sup_{\xi\in\Xi} f(x,\xi),$$

which means searching for a decision which minimizes the objective under the worst-case scenario. The solutions to this problem are usually very conservative and should be used only in specific situations. A more sophisticated approach is to aggregate the scenario specific functions into one. The most natural and common way is to use the **expected value** of the objective. If we assume that $f(x,\boldsymbol{\xi}) \in \mathcal{L}_1(\Omega,\mathcal{A},\mathcal{P}_\Omega)\ \forall x \in \mathcal{X}_0$, the new objective function can be defined as

$$\begin{aligned} F(x,\mathcal{P}) &= \mathbb{E}_{\mathcal{P}}[f(x,\boldsymbol{\xi})] \\ &= \int_\Omega f(x,\boldsymbol{\xi}(\omega))\, d\mathcal{P}_\Omega(\omega) \\ &= \int_{\mathbb{R}^d} f(x,\xi)\, d\mathcal{P}(\xi) \\ &= \int_\Xi f(x,\xi)\, d\mathcal{P}(\xi). \end{aligned}$$

The full model reformulation will then be

$$\min_{x\in\mathcal{X}_0} \mathbb{E}_{\mathcal{P}}[f(x,\boldsymbol{\xi})] = \min_{x\in\mathcal{X}_0} F(x,\mathcal{P}). \tag{1.3}$$

We will also make the assumption that the optimal solution of the problem (1.3) exists. The expected value criterion possesses a useful property, since if the functions $f(x,\xi)$ are convex on $\mathcal{X}_0$ for all $\xi \in \Xi$ , the objective function $F(x,\mathcal{P})$

is also convex on $\mathcal{X}_0$. Additionally, if the set $\mathcal{X}_0$ is convex, it results in a convex programming model.

In practice the true distribution $\mathcal{P}$ is usually not known and needs to be suitably estimated. This often results in a discrete distribution on a finite support with elements called scenarios. In other words, we consider the case $\Xi = \{\xi^1, ..., \xi^S\}$, where $S$ is a finite positive integer and the scenario probabilities are denoted $p^s$. The formulation of a stochastic optimization problem with a finite number of scenarios is

$$\min_{x \in \mathcal{X}_0} \sum_{s=1}^{S} p^s \cdot f(x, \xi^s).$$

Note that any positively weighted sum of the functions $f(x, \xi^s)$ could be used, if the decision maker has a different opinion about the importance of the scenarios. Each unique optimal solution of such a problem would be efficient with respect to the multi-objective program $\min_{x \in \mathcal{X}_0} \{f(x, \xi^1), ..., f(x, \xi^S)\}$.

### 1.2.3 Randomness in the feasibility set

Now we are looking at the case, where the objective function does not depend on the random element. The general uncertain formulation is

$$\text{"} \min_{x \in \mathcal{X}(\boldsymbol{\xi})} f(x) \text{"}$$

and the nonlinear formulation is

$$\begin{gathered} \min_{x \in \mathcal{X}_0} \ f(x) \\ s.t. \\ \text{"} g_j(x, \boldsymbol{\xi}) \leq 0\text{"}, \ j = 1, ..., p, \\ \text{"} h_k(x, \boldsymbol{\xi}) = 0\text{"}, \ k = 1, ..., q. \end{gathered}$$

The feasibility set is now random and its exact shape is unknown prior to decision making, therefore it is entirely possible to choose a solution which will turn out to be infeasible. First let us formulate the simplest method to provide some contrast. The decision maker can again use an estimate $\hat{\xi}$ of $\boldsymbol{\xi}$ and solve the problem $\min_{x \in \mathcal{X}(\hat{\xi})} f(x)$. It is a reasonable idea in the case of a random objective function, but when used for the constraints, it is far from ideal. The following motivation example demonstrates why.

*Example.* Let $x \in \mathbb{R}$ and $\boldsymbol{\xi} = (\boldsymbol{\xi_1}, \boldsymbol{\xi_2})^T$, let $\boldsymbol{\xi_1}$ have a continuous uniform distribution $\mathcal{U}[-3, -1]$, $\boldsymbol{\xi_2}$ have a continuous uniform distribution $\mathcal{U}[1, 3]$ and let them be independent. Let the objective function be $f(x) = x$ and the set of feasible solutions be determined by the random constraint "$\boldsymbol{\xi_1} \leq x \leq \boldsymbol{\xi_2}$". We can produce an estimate $\hat{\xi} = (\mathbb{E}[\boldsymbol{\xi}_1], \mathbb{E}[\boldsymbol{\xi}_2])^T = (-2, 2)^T$. The reformulated set of feasible decisions is then determined by the constraint $x \in [-2, 2]$. If we solve the problem $\min_{x \in [-2,2]} x$, we arrive at the optimal solution $x^* = -2$. However, $\mathbb{P}(\boldsymbol{\xi_1} \leq -2 \leq \boldsymbol{\xi_2}) = \frac{1}{2}$. This means that the chosen solution has a 50% chance of being infeasible with respect to the original problem.

More meaningful solutions should aim to control the probability of infeasibility. The first option is to use the robust approach again, which amounts to finding solutions from $\mathcal{X}_0$ that fulfill the random constraint at all times, i.e. $\forall \xi \in \Xi$. In other words, the solution should lie in the set $\cap_{\xi \in \Xi} \mathcal{X}(\xi)$. These are called permanently feasible solutions. The joint case for all the constraints would be to solve the program

$$\min_{x \in \cap_{\xi \in \Xi} \mathcal{X}(\xi)} f(x).$$

The severe limitation of this approach is that the set $\cap_{\xi \in \Xi} \mathcal{X}(\xi)$ is often very small or even empty. It should only be used in cases where it is absolutely necessary for the solution to remain feasible, for example when optimizing the properties of a populated building which has to withstand extreme weather in all cases. Usually there are some constraints which can be relaxed and this brings us to the second option.

The approach when the decision maker controls the probability of feasibility is called **probability constraints**. The constraints can either be controlled individually, or jointly. The two problem formulations are listed below.

**Joint probability constraints**:

$$\begin{gathered} \min_{x \in \mathcal{X}_0} \ f(x) \\ s.t. \\ \mathbb{P}(x \in \mathcal{X}(\boldsymbol{\xi})) \geq 1 - \epsilon, \end{gathered}$$

where $0 \leq \epsilon \leq 1$ is a threshold constant chosen in advance (usually close to zero).

**Individual probability constraints**

$$\begin{gathered} \min_{x \in \mathcal{X}_0} \ f(x) \\ s.t. \\ \mathbb{P}(g_j(x, \boldsymbol{\xi}) \leq 0) \geq 1 - \epsilon_j^g, \ j = 1, ..., p, \\ \mathbb{P}(h_k(x, \boldsymbol{\xi}) = 0) \geq 1 - \epsilon_k^h, \ k = 1, ..., q, \end{gathered}$$

where $\epsilon_j^g, \ j = 1, ..., p$ and $\epsilon_k^h, \ k = 1, ..., q$ are threshold constants from $[0, 1]$ chosen in advance (usually close to zero).

Note that when some threshold constant is equal to zero, it is a type of robust constraints, but we do not necessarily need the feasibility to hold for all $\xi \in \Xi$, but only on a set of probability one. These two approaches can be combined at will, the robust version can be used for more important constraints and the rest can be handled using probability constraints. There is no universal way of dealing with probability constraints reformulation, but we can list two important examples. First is when the random element can be separated from the rest of the constraint, second is when the random element has a finite number of scenarios.

*Example.* (Separated random element) Let us have a simple case of a probability constraint $\mathbb{P}(x \geq \boldsymbol{\xi}) \geq 0.95$, where $x \in \mathbb{R}$ and $\boldsymbol{\xi} \sim N(\mu, \sigma^2)$ for some known $\mu$ and $\sigma > 0$. Then we can write $\mathbb{P}(x \geq \boldsymbol{\xi}) = \mathbb{P}(\frac{x-\mu}{\sigma} \geq \frac{\boldsymbol{\xi}-\mu}{\sigma})$, where the variable $\frac{\boldsymbol{\xi}-\mu}{\sigma}$ has a standard normal distribution $N(0,1)$ with a known quantile function $q : [0,1] \to \mathbb{R}$. Then the constraint $\mathbb{P}(\frac{x-\mu}{\sigma} \geq \frac{\boldsymbol{\xi}-\mu}{\sigma}) \geq 0.95$ can be rewritten as $\frac{x-\mu}{\sigma} \geq q(0.95)$, or in other terms, $x \geq \mu + q(0.95) \cdot \sigma$, which is a deterministic inequality constraint.

*Example.* (big-$M$ reformulation) Let us have a constraint $\mathbb{P}(g(x, \boldsymbol{\xi}) \leq 0) \geq 0.95$ for $x \in \mathbb{R}^n$ and assume that the random vector $\boldsymbol{\xi}$ has a finite number of scenarios, i.e. $\Xi = \{\xi^1, ..., \xi^S\}$ with probabilities $p^s$. Then we essentially obtain $S$ scenario specific constraints $g(x, \xi^s) \leq 0,\ s = 1, ..., S$. We will create $S$ binary variables $y^s \in \{0, 1\},\ s = 1, ..., S$ which will assess whether the specific constraint is activated. The new set of $S + 1$ deterministic constraints replacing the random constraint is

$$g(x, \xi^s) \leq M \cdot (1 - y^s),\ s = 1, ..., S,$$
$$\sum_{s=1}^{S} p^s \cdot y^s \geq 0.95,$$

where $M$ is a large enough constant. We can see that if $y^s = 1$ for some scenario $s$, the constraint $g(x, \xi^s) \leq 0$ is activated, but if $y^s = 0$, we obtain $g(x, \xi^s) \leq M$ which always holds as long as we take $M \geq \sup_{x \in \mathcal{X}_0} \max_{\xi \in \Xi} g(x, \xi)$. The last constraint ensures that the desired proportion of cases is covered.

There exists a variety of assumptions which guarantee the convexity of the set of feasible solutions for probability constraints, but they are out of the scope of this thesis.

## 1.3 Multi-stage stochastic programming

The standard stochastic programming model assumes that all the decisions are made before the random element reveals itself, this is what we call a single stage model. The real life situations can be more complex than that and the randomness may reveal itself sequentially, giving the decision maker an opportunity to react in time. This is the framework of multi-stage stochastic programming. A very common special case is the two-stage stochastic programming and it is worth mentioning separately to understand the important principles. In the whole section we will assume that all expectations exist and are finite and that optimal solutions exist for each problem.

### 1.3.1 Two-stage models

Two-stage programming consists of three steps, initial decision, observation of randomness and recourse decision based on the observation. The first and third

steps are called decision stages. Let us now provide the general formulation of such a problem. Let $x \in \mathbb{R}^{n_1}$ be the first stage decision vector, $y \in \mathbb{R}^{n_2}$ the second stage decision vector and $n_1 + n_2 = n$. The random element $\boldsymbol{\xi}$ is defined the same as before. The causal pathway of the decision process is illustrated below.

$$\text{decide } x \rightsquigarrow \text{observe } \xi \rightsquigarrow \text{decide } y$$

An important observation is that the first stage decision variables are deterministic, but the second stage decision variables are dependent on the random element, therefore random. Now we are prepared to define the model.

**Definition 2.** *(Shapiro et al. (2009)) Using the notations above, we define a* ***Two-stage stochastic programming model*** *as*

$$\min_{x \in \mathcal{X}_1} \mathbb{E}_{\mathcal{P}}\Big[\min_{y \in \mathcal{X}_2(x, \boldsymbol{\xi})} f(x, y, \boldsymbol{\xi})\Big],$$

*where,* $\mathcal{X}_1 \subseteq \mathbb{R}^{n_1}$ *is a given compact set,* $\mathcal{X}_2$*:* $\mathbb{R}^{n_1} \times \mathbb{R}^d \rightrightarrows \mathbb{R}^{n_2}$ *is an* $\mathcal{A}$*- measureable closed valued multifunction and* $f : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^d \to \mathbb{R}$ *is a function.*

The widely used approach for its numerical tractability and applicability is to consider the distribution to have a finite number of scenarios, so we are again in the situation $\Xi = \{\xi^1, ..., \xi^S\}$ with probabilities $p^s$. We will create a separate second stage decision vector $y^s \in \mathbb{R}^{n_2}$ for each possible scenario $\xi^s$. The deterministic reformulation from Shapiro et al. (2009) of the two-stage scenario based stochastic program then is

$$\begin{aligned} &\min_{x, y^1, ..., y^S} \sum_{s=1}^{S} p^s \cdot f(x, y^s, \xi^s) \\ &\text{s.t. } x \in \mathcal{X}_1,\ y^s \in \mathcal{X}_2(x, \xi^s),\ s = 1, ..., S. \end{aligned}$$

The optimal solution basically serves as a decision strategy, where we take the initial optimal decision $x$ and based on the realization $\xi^s$ we choose the optimal second stage decision $y^s$. This framework gives the decision maker an ability to react optimally to the random element. What if the randomness reveals itself gradually at different points in time? It is possible to model that situation using two stages and simply wait until every realization is known to make the next move. However, a much more flexible solution would be to react to each realization in real time, prior the next realization. This way we create an entire decision process adapted to the underlying random process. This is the case of multi-stage programming.

### 1.3.2 Multi-stage models

#### General setting

As mentioned above, the decision process will consist of $T$ decision stages and $T-1$ random realizations between them, which means that the first decision vector is deterministic and the rest of the decision process is random. Note that literature sometimes states the case, where there is one additional random element taking place after the last decision, with no recourse action. To specify further, let $n_1, ..., n_T, d_1, ..., d_{T-1}$ be positive integers, $n = n_1 + ... + n_T$ and $d = d_1 + ... + d_{T-1}$. The vector of all decisions will be denoted $x \in \mathbb{R}^n$ such that $x = (x_1, ..., x_T)^T$, where $x_t \in \mathbb{R}^{n_t}$, $t \in \{1, ..., T\}$ is the decision vector in the $t$-th stage of the program. By $x_{[t]}$ we will denote all the decisions made until the $t$-th stage, i.e. $x_{[t]} = (x_1, ..., x_t)^T$.

The random element in the problem will be represented by a real stochastic process $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_{T-1})^T$, where for $t \in \{1, ..., T-1\}$ we have a random vector $\boldsymbol{\xi}_t : (\Omega, \mathcal{A}) \to (\mathbb{R}^{d_t}, \mathcal{B}^{d_t})$ encapsulating the random element revealed to the decision maker after the $t$-th stage decision. Same as for the decision process, we denote by $\boldsymbol{\xi}_{[t]} = (\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_t)^T$ the history of the process up to time $t$ and by $\mathcal{Z}_t \subseteq \mathcal{A}$ the $\sigma$-algebra generated by the process up to time $t$. $\{\mathcal{Z}_t\}_{t=1}^{T-1}$ is then the canonical filtration of the process. By $\boldsymbol{\xi}_{[0]}$ we mean that no random information has been revealed yet. The causal pathway of this model is illustrated below.

$$\text{decide } x_1 \rightsquigarrow \text{observe } \xi_1 \rightsquigarrow \text{decide } x_2 \rightsquigarrow ... \rightsquigarrow \text{observe } \xi_{T-1} \rightsquigarrow \text{decide } x_T$$

*Remark.* (Shapiro et al. (2009)) We can pose additional assumptions about the underlying stochastic process. For example, we call it **stagewise independent** when $\boldsymbol{\xi}_t$ is stochastically independent of $\boldsymbol{\xi}_{[t-1]}$, or **Markovian** when the conditional distribution of $\boldsymbol{\xi}_t | \boldsymbol{\xi}_{[t-1]}$ is the same as for $\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t-1}$.

#### Model definition

Now we can present the general form of a multi-stage stochastic program.

**Definition 3.** *(Shapiro et al. (2009)) Using the notations above, we define a* ***Multi-stage stochastic programming model*** *as*

$$\min_{x_1 \in \mathcal{X}_1} f_1(x_1) + \mathbb{E}\Big[\min_{x_2 \in \mathcal{X}_2(x_1, \boldsymbol{\xi}_1)} f_2(x_2, \boldsymbol{\xi}_1) + \mathbb{E}\Big[... + \mathbb{E}\Big[\min_{x_T \in \mathcal{X}_T(x_{T-1}, \boldsymbol{\xi}_{T-1})} f_T(x_T, \boldsymbol{\xi}_{T-1})\Big]\Big]\Big],$$

*where $f_1 : \mathbb{R}^{n_1} \to \mathbb{R}$ is a deterministic function, $\mathcal{X}_1 \subseteq \mathbb{R}^{n_1}$ is a fixed compact set, for $t = 2, ..., T$ are $f_t : \mathbb{R}^{n_t} \times \mathbb{R}^{d_{t-1}} \to \mathbb{R}$ functions and $\mathcal{X}_t : \mathbb{R}^{n_{t-1}} \times \mathbb{R}^{d_{t-1}} \rightrightarrows \mathbb{R}^{n_t}$ $\mathcal{A}$-measureable closed-valued multifunctions. The expectation in the $t$-th stage is taken with respect to the conditional distribution of $\boldsymbol{\xi}_t | \boldsymbol{\xi}_{[t-1]}$, $t = 1, ..., T-1$.*

In the case where some of the optimal solutions don't exist, we could replace the minima by infima.

**Scenario trees**

The next two sections are partly drawn from Shapiro et al. (2009). The single stage models did not include any possibilities for recourse action. Because of that, it was simple to capture the scenario structure, we just approximated the distribution $\mathcal{P}$ by a finite number of atoms. This situation is more delicate, since we are dealing with a random process. We can again assume that the element $\boldsymbol{\xi}$ has finitely many scenarios $\xi^1, ..., \xi^S$, but we have to realize, that each of these scenarios is a whole trajectory of the underlying random process, i.e. $\xi^s = (\xi_1^s, ..., \xi_{T-1}^s)^T$, where the $t$-th element is a vector from $\mathbb{R}^{d_t}$. We can think of every scenario $\xi^s$ as a path from a root node to a leaf node in a scenario tree. The nodes of the tree are organized in levels corresponding to the decision stages $t = 1, ..., T$. Each node corresponds to a specific situation the decision maker might find themselves in. At level $t = 1$ we have one root node, marking the beginning of the decision process, and at level $t = T$ we have S leaf nodes, each corresponding to one possible trajectory $\xi^s$. Nodes in the tree are connected via arcs to nodes in the next level, signifying a causal pathway between them, defined by the realizations of the process. We denote $N_t$ for $t = 1, ..., T$ the sets of nodes in each level $t$. Each node $\nu \in N_t$ for $t = 2, ..., T$ is connected to exactly one node in the previous level, called the ancestor node $a(\nu) \in N_{t-1}$ and each node $\nu \in N_t$ for $t = 1, ..., T-1$ has a set of descendants $d(\nu) \subseteq N_{t+1}$ in the next level, which all correspond to different continuations of the nodes history $\xi_{[t-1]}$. It is a simple observation that the set of nodes $N_{t+1}$ is a disjoint union of descendant sets from the previous layer, i.e. $N_{t+1} = \cup_{\nu \in N_t} d(\nu)$, and $d(\nu_1) \cap d(\nu_2) = \emptyset$ if $\nu_1 \neq \nu_2$. In Figure 1.1 we can see an illustration of how such a scenario tree would look like in a three-stage stagewise independent stochastic program, where $\boldsymbol{\xi_1}$ has 3 possible realizations and $\boldsymbol{\xi_2}$ has 2 possible realizations, leading to 6 different trajectory scenarios $\xi^1, ..., \xi^6$ of the process $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)^T$ in the last level. In the figure we can also see one specific node $\nu$ in the second layer and his ancestor and two descendants.

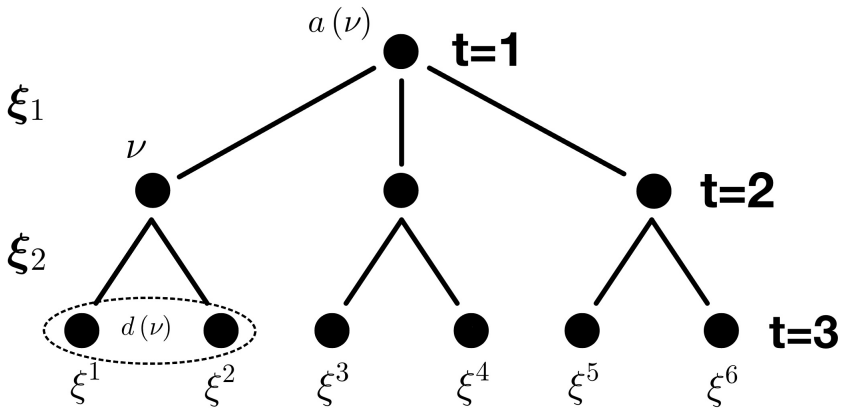


Figure 1.1: Example of a scenario tree for a three-stage stochastic program

After properly identifying the tree structure, the probability distribution needs to be assessed. In order to do that, we have to specify the conditional distributions

of $\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_{[t]}$, $t = 1, ..., T-2$. Generally, for an arbitrary node $\nu \in N_t$ at a time $t = 1, ..., T-1$, it is necessary to specify the probabilities $p_{\nu,\eta} \geq 0$ of moving to different descendants $\eta \in d(\nu)$, such that $\sum_{\eta \in d(\nu)} p_{\nu,\eta} = 1$. These probabilities are in one-to-one correspondence with the arcs inside the tree and they represent the conditional distributions of the process continuations. Each trajectory $\xi^s$ corresponds to a chronological sequence of nodes $\nu_1, ..., \nu_T$, such that $\nu_t \in N_t$ and $a(\nu_t) = \nu_{t-1}$. Utilizing that $N_1 = \{\nu_1\}$, the probability of the scenario $\xi^s$ is then equal to $p^s = p_{\nu_1,\nu_2} \cdot p_{\nu_2,\nu_3} \cdot ... \cdot p_{\nu_{T-1},\nu_T}$. Earlier on we mentioned possible assumptions about the underlying process, such as stagewise independence or Markovian property. These can make the probabilistic structure simpler. For example, in the former case, we can calculate a nodes probability as a product of marginal probabilities of the events leading up to that node. In the Markovian case, we can calculate it by utilizing one step transition probabilities from the ancestral node.

### Nonanticipativity

Remember that for each trajectory scenario $\xi^s = (\xi_1^s, ..., \xi_{T-1}^s)^T$ there is a corresponding decision process $x^s = (x_1^s, ..., x_T^s)^T$. It is extremely important to realize, that until the last stage, the decision maker does not know in which scenario $\xi^s$ they are, since they only know the already observed part of the process. This is why the model formulation needs to ensure that decisions must not be made based on future realizations of the random process. In other words, we need the decision $x_t$ to be $\mathcal{Z}_{t-1}$-measureable. The nested multi-stage formulation in Definition 3 included this principle implicitly, but it needs to be manually incorporated into the deterministic reformulations via so called **nonanticipativity constraints**. Specifically, let us look at Figure 1.1 one more time, and imagine we are now at the second stage in the rightmost branch of the tree, waiting to make our second stage decision. However, we have no way of knowing if we will end up in scenario $\xi^5$ or $\xi^6$, since it depends on the realization of the random vector $\boldsymbol{\xi_2}$, which has not been revealed yet. In order to have a correctly adapted decision process, we need to ensure that scenarios $\xi^5$ and $\xi^6$ have the same decision policy up to the second stage, since they share a common history. We need to look at each pair of scenarios and compare their histories. Mathematically, for each stage $t = 1, ..., T$ and each pair of scenarios $s_1, s_2$, if $\xi_{[t-1]}^{s_1} = \xi_{[t-1]}^{s_2}$, we enforce a constraint $x_t^{s_1} = x_t^{s_2}$. A simple consequence is, that the root node is the same for all scenarios, meaning that it must hold $x_1^1 = x_1^2 = ... = x_1^S$, i.e. the first stage decision is the same no matter the scenario, since there is no randomness revealed yet.

# 2. Endogenous randomness in stochastic programming

In the beginning of the first chapter we stated a core assumption of classical stochastic programming, that the random element is not affected by the decision maker and its distribution is fixed. This situation was referred to as exogenous randomness, indicating that the random element is somehow external to the decision process. There are actually many cases, where this assumption does not hold. It is difficult to generalize them, but they usually arise when the randomness is somehow coming from within the model framework, or at least partly. This is why it is called endogenous randomness. Sometimes the term decision dependent randomness will be used in the same meaning. We will see several examples of such situations, but for a quick introduction, imagine a salesman pitching a new product and deciding on the selling price. The demand for this product is surely random, however, is it reasonable to assume that its distribution does not depend on the price? We have to keep in mind, that incorporating endogenous randomness into stochastic programming makes the models more complex and computationally expensive, especially in multi-stage programming, where the decisions may even change the structure of the underlying random process and the times at which the realizations become known. This is sometimes called endogenous randomness of the second type and it is a very complex topic, which will not be covered in sufficient detail.

There is vast amount of possible ways how a variable can affect a probability distribution and this thesis can not provide an exhaustive list of them. The goal will be to search for tractable cases, which are suitable for real life applications. The search for tractable cases is not only important from a modelling standpoint, but also from a computational standpoint, as some models with endogenous randomness can be extremely hard to solve or even formulate. Let us now lay out some basic mathematical background, in which they can be formulated.

## 2.1 General setting

Let us consider the previously defined probability space $(\Omega, \mathcal{A}, \mathcal{P}_\Omega)$ and closed set of feasible decisions $\mathcal{X}_0 \subseteq \mathbb{R}^n$. Let us denote by $\mathcal{P}_0$ an arbitrary set of distributions on the measurable space $(\mathbb{R}^d, \mathcal{B}^d)$. It could include all distributions on the measurable space, but it also allows us to restrict to some specific subset. For example, we might only want to include distributions with a finite expectation. For now we will assume that every feasible decision yields exactly one probability distribution, i.e. for every feasible decision $x \in \mathcal{X}_0$ is the underlying random element represented by a random vector $\boldsymbol{\xi}(x) : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}^d)$ with a support $\Xi(x) \subseteq \mathbb{R}^d$ and a probability distribution $\mathcal{P}(x) \in \mathcal{P}_0$. The distribution function of $\boldsymbol{\xi}(x)$ will be denoted $\Delta(x, \cdot) : \mathbb{R}^d \to [0, 1]$ and the density $\delta(x, \cdot) : \mathbb{R}^d \to [0, \infty)$. We also denote $\mathcal{P}(\mathcal{X}_0) = \{\mathcal{P}(x) : \ x \in \mathcal{X}_0\} \subseteq \mathcal{P}_0$ the set of all distributions which

can be determined by feasible decisions. Note that the random elements do not need to have the same dimension for every decision, we can choose the constant $d$ as the maximal dimension of the random elements from $\{\boldsymbol{\xi}(x) : x \in \mathcal{X}_0\}$ and the framework allows for some of the random variables inside each vector to degenerate to zero or some other constant. In order to not have to segregate the random variables into exogenous and endogenous, some of the marginal distributions of $\boldsymbol{\xi}(x)$ can be independent of decisions (possibly even all of them, which would simply result in the classical stochastic programming framework).

### 2.1.1 Model formulation

The general uncertain form of a stochastic programming model with endogenous randomness is

$$\begin{aligned} &\text{"}\min_{x\in\mathcal{X}_0} \ f(x, \boldsymbol{\xi}(x))\text{"} \\ &\quad s.t. \ (x, \boldsymbol{\xi}(x)) \in \Psi, \end{aligned}$$

where $f : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ is a function and the set $\Psi$ expresses the random constraints. In order to be consistent with the previous chapter, we will stay in the framework of nonlinear optimization, where the set of random constraints consists of a finite number of equalities and inequalities.

**Definition 4.** *We define a* ***Stochastic programming model with endogenous randomness*** *as*

$$\begin{aligned} &\text{"}\min_{x\in\mathcal{X}_0} \ f(x, \boldsymbol{\xi}(x))\text{"} \\ &\qquad s.t. \\ &\text{"}g_j(x, \boldsymbol{\xi}(x)) \leq 0\text{"}, \ j = 1, ..., p, \\ &\text{"}h_k(x, \boldsymbol{\xi}(x)) = 0\text{"}, \ k = 1, ..., q, \end{aligned} \tag{2.1}$$

*where all the functions are defined the same as in the corresponding exogenous formulation.*

*Remark.* In the above case it holds that

$$\begin{aligned} \Psi = \{(x, \boldsymbol{\xi}(x)) : &\text{"}g_j(x, \boldsymbol{\xi}(x)) \leq 0\text{"}, \ j = 1, ..., p, \\ &\text{"}h_k(x, \boldsymbol{\xi}(x)) = 0\text{"}, \ k = 1, ..., q\}. \end{aligned}$$

Same as before, this is the uncertain form of the model, which needs to be reformulated deterministically. The tractability of a stochastic programming model with endogenous randomness depends a lot on its specific structure and searching for such structures is still an ongoing process in literature.

### 2.1.2 Deterministic reformulations

For simplicity, we will now assume a fixed feasibility set. In other words, we are dealing with the problem

$$\text{"}\min_{x\in\mathcal{X}_0} f(x,\boldsymbol{\xi}(x))\text{"}.$$

First option is to use the robust **worst-case** formulation

$$\min_{x\in\mathcal{X}_0}\sup_{\xi\in\Xi(x)} f(x,\xi),$$

which is much more complicated than the exogenous version, because the supports $\Xi(x)$ are decision dependent and can differ for each $x$. This makes the model viable only when the support dependence is simple, for example when the decisions translate a common support by some decision dependent constant.

The standard approach is to use the **expected value criterion**, same as before. Assume for each $x \in \mathcal{X}_0$ that $f(x,\boldsymbol{\xi}(x)) \in \mathcal{L}_1(\Omega,\mathcal{A},\mathcal{P}_\Omega)$. We can define a function $F : \mathcal{X}_0 \times \mathcal{P}_0 \to \mathbb{R}$ by the following expressions.

$$\begin{aligned} F(x,\mathcal{P}(x)) &= \mathbb{E}_{\mathcal{P}(x)}[f(x,\boldsymbol{\xi}(x))] \\ &= \int_\Omega f(x,\boldsymbol{\xi}(x)(\omega))\, d\mathcal{P}_\Omega(\omega) \\ &= \int_{\mathbb{R}^d} f(x,\xi)\, d\mathcal{P}(x)(\xi) \\ &= \int_{\Xi(x)} f(x,\xi)\, d\mathcal{P}(x)(\xi). \end{aligned} \tag{2.2}$$

The full model reformulation is

$$\min_{x\in\mathcal{X}_0} F(x,\mathcal{P}(x)) = \min_{x\in\mathcal{X}_0} \mathbb{E}_{\mathcal{P}(x)}[f(x,\boldsymbol{\xi}(x))]. \tag{2.3}$$

Remember that in the exogenous case, convexity of the function $f$ implicated the convexity of $F(x,\mathcal{P})$. Dupačová (2006) mentioned that this property may be lost in the endogenous case. This provides further limitations in the search for tractable cases.

At this point we can not provide a general reformulation of the random constraints, instead we will search for specific model structures where the reformulation can be done in a more general manner. Let us now proceed to cover some possible cases of decision dependence. Whatever reformulation of the general model (2.1) we will consider in the future, we will always assume the existence of an optimal solution. From now on, the assumption that $f(x,\boldsymbol{\xi}(x)) \in \mathcal{L}_1(\Omega,\mathcal{A},\mathcal{P}_\Omega)$ holds for each $x \in \mathcal{X}_0$ will be implicitly considered, even when not explicitly stated. Moreover, we will work with the expected value criterion $F(x,\mathcal{P}(x))$ defined in (2.2) as the objective function, unless stated otherwise.

## 2.2 Common reference measure

Let us now look at one possible simplification of the general model with a fixed feasibility set. As proposed in Dupačová (2006), assume that all the probability measures $\mathcal{P}(x) \in \mathcal{P}(\mathcal{X}_0)$ have existing densities $\delta(x, \cdot) : \mathbb{R}^d \to [0, \infty)$ with respect to a common probability measure $\mathcal{G}$. The objective function can be rewritten as

$$F(x, \mathcal{P}(x)) = \int_{\mathbb{R}^d} f(x, \xi)\, d\mathcal{P}(x)(\xi) = \int_{\mathbb{R}^d} f(x, \xi) \cdot \delta(x, \xi)\, d\mathcal{G}(\xi)$$

and by defining $\tilde{f}(x, \xi) = f(x, \xi) \cdot \delta(x, \xi)$ we can write the whole problem as

$$\min_{x \in \mathcal{X}_0} \int_{\mathbb{R}^d} \tilde{f}(x, \xi)\, d\mathcal{G}(\xi)$$

with a decision independent distribution $\mathcal{G}$. Unfortunately, this can lead to a loss of convenient properties of the original objective function $f$ (especially convexity), which makes the evaluation of gradients and subgradients more challenging. The structure of the problem is of high importance here. For example, we can formulate a special case of the dependence is proposed by Dupačová (2006): $\mathcal{P}(x)(B) = \mathcal{G}(B \bigoplus Hx)$, where $B \in \mathcal{B}^d$ is a Borel set, $\bigoplus$ is the direct sum and H is a matrix of corresponding dimension. This situation yields an objective function

$$F(x, \mathcal{P}(x)) = \int_{\mathbb{R}^d} f(x, \zeta - Hx)\, d\mathcal{G}(\zeta),$$

which retains the properties of the original objective function $f$.

As for the random constraints, we can formulate a special case proposed by Kopa (2024). Let the random part of the feasibility set consist of constraints in the shape $\mathbb{E}_{\mathcal{P}(x)}[g_j(x, \boldsymbol{\xi}(x))] \leq 0$ for each $j = 1, ..., p$. These can be equivalently written as one constraint $\max_{j=1,...,p} \mathbb{E}_{\mathcal{P}(x)}[g_j(x, \boldsymbol{\xi}(x))] \leq 0$. If we can find a function $l(x, \boldsymbol{\xi}(x))$ such that

$$\max_{j=1,...,p} \mathbb{E}_{\mathcal{P}(x)}[g_j(x, \boldsymbol{\xi}(x))] = \mathbb{E}_{\mathcal{P}(x)}[l(x, \boldsymbol{\xi}(x))],$$

the whole problem (2.1) can be reformulated as

$$\begin{aligned} \min_{x \in \mathcal{X}_0} \quad & \int_{\mathbb{R}^d} f(x, \xi) \cdot \delta(x, \xi)\, d\mathcal{G}(\xi) \\ s.t. \quad & \int_{\mathbb{R}^d} l(x, \xi) \cdot \delta(x, \xi)\, d\mathcal{G}(\xi) \leq 0. \end{aligned}$$

## 2.3 Finite cardinality of $\mathcal{P}(\mathcal{X}_0)$

This type of decision dependence is discussed in detail in Kopa (2024). Let there be an integer $m \geq 1$ and random vectors $\boldsymbol{\xi_1}, ..., \boldsymbol{\xi_m} : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathbb{B}^d)$ with distributions $\mathcal{P}_1, ..., \mathcal{P}_m \in \mathcal{P}_0$. Let there be a finite disjoint partition of the feasibility set $\mathcal{X}_0$, such that each part has a common random element. In mathematical terms, there is a disjoint partition $\mathcal{X}_0 = \cup_{i=1}^m \mathcal{X}_i$, such that $\forall i = 1, ..., m$ and $\forall x \in \mathcal{X}_i : \boldsymbol{\xi}(x) \overset{\text{a.s.}}{=} \boldsymbol{\xi_i}$ and $\mathcal{P}(x) = \mathcal{P}_i$. The simple observation is that $\mathcal{P}(\mathcal{X}_0) = \{\mathcal{P}_1, ..., \mathcal{P}_m\}$. Assume that for each $i = 1, ..., m$ is $F(x, \mathcal{P}_i)$ continuous in $x$ on the set $clo(\mathcal{X}_i)$, where $F$ is defined in (2.2). This way we can split the original problem into $m$ subproblems, which all follow the classical stochastic programming framework with exogenous randomness. For each $i = 1, ..., m$, the uncertain form of the $i$-th subproblem is

$$
\begin{gathered}
\text{"} \min_{x \in clo(\mathcal{X}_i)} \; f(x, \boldsymbol{\xi_i})\text{"} \\
s.t. \\
\text{"} g_j(x, \boldsymbol{\xi_i}) \leq 0\text{"}, \; j = 1, ..., p, \\
\text{"} h_k(x, \boldsymbol{\xi_i}) = 0\text{"}, \; k = 1, ..., q.
\end{gathered}
$$

Firstly, let us consider a fixed feasibility set. The subproblem for each $i = 1, ..., m$ can be reformulated using the expected value criterion into the shape

$$
\min_{x \in clo(\mathcal{X}_i)} F(x, \mathcal{P}_i) = \min_{x \in clo(\mathcal{X}_i)} \mathbb{E}_{\mathcal{P}_i}[f(x, \boldsymbol{\xi_i})].
$$

After solving all the subproblems, we can create an index set $I^* \subseteq \{1, ..., m\}$, indicating which of the subproblems have a solution. This set will always be non-empty, since we consider the whole problem to have a solution. The reformulated model (2.3) with a fixed feasibility set is

$$
\min_{i \in I^*} \min_{x \in clo(\mathcal{X}_i)} F(x, \mathcal{P}_i) = \min_{i \in I^*} \min_{x \in clo(\mathcal{X}_i)} \mathbb{E}_{\mathcal{P}_i}[f(x, \boldsymbol{\xi_i})].
$$

The principle stays the same when we allow random constraints. For each of the subproblems there is a common distribution $\mathcal{P}_i$, therefore we can use probability constraints defined within the exogenous framework. Utilizing the same definition of $I^*$, a reformulation of the full model (2.1) with individual probability constraints is

$$\min_{i \in I^*} \Phi(\mathcal{P}_i),$$

where

$$\Phi(\mathcal{P}_i) = \min_{x \in clo(\mathcal{X}_i)} \mathbb{E}_{\mathcal{P}_i}[f(x, \boldsymbol{\xi_i})]$$
$$s.t.$$
$$\mathbb{P}(g_j(x, \boldsymbol{\xi_i}) \leq 0) \geq 1 - \epsilon_j^g, \ j = 1, ..., p,$$
$$\mathbb{P}(h_k(x, \boldsymbol{\xi_i}) = 0) \geq 1 - \epsilon_k^h, \ k = 1, ..., q.$$

The submodels could alternatively be formulated using joint probability constraints in the form

$$\mathbb{P}\Big(g_j(x, \boldsymbol{\xi_i}) \leq 0, \ j = 1, ..., p, \ h_k(x, \boldsymbol{\xi_i}) = 0, \ k = 1, ..., q\Big) \geq 1 - \epsilon.$$

*Remark.* Note that it would also be possible to model this situation using binary variables by defining $y_1, ..., y_m \in \{0, 1\}$ such that $y_i = 1 \Leftrightarrow x \in \mathcal{X}_i$, the problem with fixed feasibility set could be formulated as

$$\min_{x \in \mathcal{X}_0, \ y_1, ..., y_m \in \{0,1\}} \sum_{i=1}^{m} y_i \cdot \mathbb{E}_{\mathcal{P}_i}[f(x, \boldsymbol{\xi_i})]$$
$$s.t. \ y_i = 1 \Leftrightarrow x \in \mathcal{X}_i, \ i = 1, ..., m.$$

## 2.4 Fixed parametric family

Let us consider an arbitrary parametric family of distributions $\{\mathcal{F}(\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ where $\boldsymbol{\theta} \in \mathbb{R}^p$ is the parameter vector and $\boldsymbol{\Theta} \subseteq \mathbb{R}^p$ is the parametric space.

*Example.* Some of the well known parametric families are:

- $\{N_d(\mu, \Sigma) \mid \mu \in \mathbb{R}^d, \ \Sigma \in \mathbb{R}^{d \times d}$ positive semi-definite$\}$ - Normal distribution of general dimension $d \in \mathbb{N}$
- $\{Exp(\lambda) \mid \lambda \in (0, \infty)\}$ - Exponential distribution
- $\{Alt(p) \mid p \in [0, 1]\}$ - Alternative or Bernoulli distribution

A very practical approach to endogenous randomness modelling is to keep the possible distributions inside a fixed parametric family. In such a case, we assume that the decisions only alter parameters, but not the type of the distribution. Mathematically, let us consider a parametric family which fits into the previously set space of distributions $\mathcal{P}_0$, i.e. $\mathcal{F} = \{\mathcal{F}(\boldsymbol{\theta}) \mid \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p\} \subseteq \mathcal{P}_0$. For each distribution from the family, corresponding to the parameter vector $\boldsymbol{\theta}$, we denote by $\Delta(\boldsymbol{\theta}, \cdot) : \mathbb{R}^d \to [0, 1]$ the cumulative distribution function and assume existence of a density $\delta(\boldsymbol{\theta}, \cdot) : \mathbb{R}^d \to [0, \infty)$ with respect to a measure $\mathcal{G}$. Then let us assume the existence of a parametric function $\theta : \mathcal{X}_0 \to \boldsymbol{\Theta}$ which serves as a mapping

between the feasible decision space and parametric space. Then for each feasible decision $x \in \mathcal{X}_0$ we assume that the distribution of $\boldsymbol{\xi}(x)$ is

$$\mathcal{P}(x) = \mathcal{F}(\theta(x)).$$

Assuming a fixed feasibility set, we can utilize this shape of dependence and existence of densities to rewrite the objective function as

$$F(x, \mathcal{P}(x)) = \int_{\mathbb{R}^d} f(x, \xi)\, d\mathcal{P}(x)(\xi) = \int_{\mathbb{R}^d} f(x, \xi) \cdot \delta(\theta(x), \xi)\, d\mathcal{G}(\xi),$$

therefore the full problem is

$$\min_{x \in \mathcal{X}_0} \int_{\mathbb{R}^d} f(x, \xi) \cdot \delta(\theta(x), \xi)\, d\mathcal{G}(\xi).$$

*Example.* Say there is a function $\lambda : \mathcal{X}_0 \to (0, \infty)$ and for each feasible decision $x$ the random element $\boldsymbol{\xi}(x)$ is univariate and follows an exponential distribution with parameter $\lambda(x)$, thus having a density with respect to the Lebesgue measure and a support $\Xi(x) = (0, \infty)$. The endogenous randomness stochastic problem for a fixed feasibility set and an objective function $f(x, \boldsymbol{\xi}(x))$ can therefore be formulated as

$$\min_{x \in \mathcal{X}_0} \int_0^\infty f(x, \xi) \cdot \lambda(x) \cdot e^{-\lambda(x) \cdot \xi}\, d\xi.$$

As for the case of a random feasibility set, we will assume only inequality constraints in the uncertain form "$g_j(x) \geq \boldsymbol{\xi}(x)$" for all $j = 1, ..., p$. Assume that $\boldsymbol{\xi}(x)$ has a distribution $\mathcal{F}(\theta(x))$ with a distribution function $\Delta(\theta(x), \cdot)$ and a quantile function $\Delta^{-1}(\theta(x), \cdot)$. Then we can, for example, use individual probability constraints for each $j = 1, ..., p$ in the form

$$\begin{aligned} &\mathbb{P}(g_j(x) \geq \boldsymbol{\xi}(x)) \geq 1 - \epsilon \\ &\Delta(\theta(x), g_j(x)) \geq 1 - \epsilon \\ &g_j(x) \geq \Delta^{-1}(\theta(x), 1 - \epsilon), \end{aligned}$$

which is a well defined deterministic constraint. Below we show a simple example for one probability constraint with a separated random element.

*Example.* Let us consider a simple case of a probability constraint in the form $\mathbb{P}(g_1(x) \geq \boldsymbol{\xi}(x)) \geq 0.95$, where $x \in \mathbb{R}$ and for every feasible $x \in \mathcal{X}_0$ does $\boldsymbol{\xi}(x)$ have a univariate normal distribution with the mean $\mu(x)$ and variance $\sigma^2(x)$ for some correctly defined parametric mappings $\mu, \sigma$. Assume that $\sigma(x) > 0$ for all feasible solutions. Then we can write $\mathbb{P}(g_1(x) \geq \boldsymbol{\xi}(x)) = \mathbb{P}\left(\frac{g_1(x) - \mu(x)}{\sigma(x)} \geq \frac{\boldsymbol{\xi}(x) - \mu(x)}{\sigma(x)}\right)$, where the variable $\frac{\boldsymbol{\xi}(x) - \mu(x)}{\sigma(x)}$ has a standard normal distribution $N(0, 1)$ with a known quantile function $q : [0, 1] \to \mathbb{R}$. Then the constraint

$\mathbb{P}\left(\frac{g_1(x)-\mu(x)}{\sigma(x)} \geq \frac{\boldsymbol{\xi}(x)-\mu(x)}{\sigma(x)}\right) \geq 0.95$ can be rewritten as $\frac{g_1(x)-\mu(x)}{\sigma(x)} \geq q(0.95)$, or in other terms, $g_1(x) \geq \mu(x) + q(0.95) \cdot \sigma(x)$, which is a deterministic inequality constraint depending only on x.

*Example.* (Motivation) Say there is a farmer who wants to maximize the yield of an arbitrary crop after planting $n \in \mathbb{N}$ seeds (each seed can yield one crop). The random element entering the model is the number of crops, which can be modeled by a binomial distribution with $n$ trials and a probability of yield for one seed. There are two influential non-negative continuous decision variables, the amount of fertilizer used $x_1$ and the amount of water for irrigation $x_2$, denote $x = (x_1, x_2)^T$. These factors influence the probability of the seed yielding a crop. Say there is a function $p : [0, \infty)^2 \to [0, 1]$, which takes fertilization and irrigation as inputs and determines the probability of yield. The goal could be to work with a fixed budget and maximize the expected profit, computed as the gain from crop sales minus the cost of fertilization and irrigation. The decision dependent random element is the crop amount $\boldsymbol{\xi}(x)$ with a distribution $\mathcal{P} = Bi(n, p(x))$.

## 2.5 Fixed and finite number of scenarios

Now we can move to arguably the most important case of endogenous randomness modelling, which is widely used in practice. We already mentioned in the first chapter, that an efficient way to work with a probability distribution is to assume it has finitely many scenarios. This discretization allows us to formulate difficult problems elegantly, without having to deal with complex shapes of the objective function, and it can preserve useful properties of the function $f$. The reason why this is a frequently used model is, that there are several situation in practise, where me might need it. First situation is, when the true distribution itself is discrete with finitely many scenarios. Second, when it is not discrete but can be approximated by discrete using various methods for scenario generation. Third but not last, when we observe a random sample from the true distribution and work with sample estimates of the expected value and other quantities, then we use arguments like the law of large numbers or central limit theorem to justify it.

### 2.5.1 Decision dependent probabilities

Either way, let us now assume that all the possible realizations of all the random elements $\boldsymbol{\xi}(x), x \in \mathcal{X}_0$ form a set of finite cardinality, i.e.

$$\Xi = \cup_{x \in \mathcal{X}_0} \Xi(x) = \{\xi^1, ..., \xi^S\}$$

for some integer $S < \infty$. Under this assumption, all distributions from $\mathcal{P}(\mathcal{X}_0)$ can be represented by a vector of scenario probabilities, which can be decision dependent. In other words, for each $x \in \mathcal{X}_0$ is the distribution $\mathcal{P}(x)$ discrete on the set $\{\xi^1, ..., \xi^S\}$ with probabilities $p^1(x), ..., p^S(x)$, which satisfy the conditions

$p^s(x) \geq 0 \ \forall s = 1, ..., S$ and $\sum_{s=1}^{S} p^s(x) = 1$. Note that we purposely did not restrict the probabilities from being zero, in order for this framework to allow the distributions to degenerate to a smaller number of scenarios, but all in the set $\Xi$. The reformulated problem (2.3) with a fixed feasibility set is

$$\min_{x \in \mathcal{X}_0} \sum_{s=1}^{S} p^s(x) \cdot f(x, \xi^s).$$

Additional assumptions can be posed on the probability functions, like continuity. An example of the use of this modelling technique can be seen in the paper Kopa and Rusý (2021), where one of the random elements was whether a client accepts a loan from a bank or not. This element obviously had two scenarios, and their probabilities were dependent on the offered interest rate, which was one of the decision variables. In this case of two scenarios, the probabilities can be modeled by logistic regression for example, with the decision variables as predictors and regression parameters estimated from data.

As for the random feasibility set, the probability constraints approach is available. One possibility is to use the aforementioned big-$M$ reformulation. Say we have a constraint $\mathbb{P}(g(x, \boldsymbol{\xi}(x)) \leq 0) \geq 1 - \epsilon$. We utilize the fact, that the scenario set is independent of decisions, and for each $s = 1, ..., S$ we create a constraint $g(x, \xi^s) \leq 0$ and a binary variable $y^s \in \{0, 1\}$, which will assess, whether this specific constraint is activated. Then we add the final constraint ensuring that the desired probability is covered. The new set of $S + 1$ deterministic constraints then is

$$\begin{aligned} & g(x, \xi^s) \leq M \cdot (1 - y^s), \ s = 1, ..., S, \\ & \sum_{s=1}^{S} p^s(x) \cdot y^s \geq 1 - \epsilon, \end{aligned}$$

where $M \geq \sup_{x \in \mathcal{X}_0} \max_{\xi \in \Xi} g(x, \xi)$. The principle is the same as before, only now are the probabilities also functions of $x$.

### 2.5.2 Decision dependent scenarios

The last model assumed fixed scenarios and decision dependent probabilities. This situation can also be considered in reverse. Let $S$ be defined the same as before. Assume that for each $x \in \mathcal{X}_0$ is the distribution $\mathcal{P}(x)$ concentrated on $S$ different decision dependent scenarios $\Xi(x) = \{\xi^1(x), ..., \xi^S(x)\}$ with fixed probabilities $p^1, ..., p^S > 0$ such that $\sum_{s=1}^{S} p^s = 1$. In other words, it is the opposite of the last situation, the probabilities are fixed and scenarios are decision dependent. Here we assume the probabilities to be non-zero, because we do not want the distributions to degenerate. We also assume that for each feasible $x$ and two different scenarios $s_1, s_2 = 1, ..., S$ it holds that $\xi^{s_1}(x) \neq \xi^{s_2}(x)$. Under these circumstanes, we can reformulate the problem (2.3) with a fixed feasibility set as

$$\min_{x \in \mathcal{X}_0} \sum_{s=1}^{S} p^s \cdot f(x, \xi^s(x)).$$

We utilize the fact, that even though the scenarios are decision dependent, there is a fixed number of them, otherwise the formulation would be complicated. This way we can use the big-$M$ reformulation again. Each probability constraint $\mathbb{P}(g(x, \boldsymbol{\xi}(x)) \leq 0) \geq 1 - \epsilon$ can be replaced by $S + 1$ new constraints

$$\begin{aligned} &g(x, \xi^s(x)) \leq M \cdot (1 - y^s), \;\; s = 1, ..., S, \\ &\sum_{s=1}^{S} p^s \cdot y^s \geq 1 - \epsilon, \end{aligned}$$

where $M \geq \sup_{x \in \mathcal{X}_0} \max_{\xi \in \Xi(x)} g(x, \xi)$. This method can be used efficiently when the dependence between the decisions and scenarios is simple. One option could be that the decisions translate the scenarios by some constant.

## 2.6 Effect on dependence structure

So far we have considered cases, where the decisions influence the whole joint distribution of the random element. What if we wanted the decisions to only alter the marginal distributions, but not the dependence between them? For that we need a mathematical tool which allows us to inspect these two concepts separately, and that tool is called a **copula**. Unlike the classical measures of dependence, mainly correlation, which only measures the rate of linear dependence, copulas allow for flexible modelling of a wide variety of possible situations. In finance for example, there can be two assets, whose returns are generally weakly correlated, but they exhibit a similar behaviour in the most extreme cases.

### 2.6.1 Theoretical basics of copulas

Let us first define the notion of a copula. Assume we have real random variables $\boldsymbol{\xi_1}, ..., \boldsymbol{\xi_d} : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ with **continuous and strictly increasing** cumulative distribution functions $\Delta_1, ..., \Delta_d : \mathbb{R} \to [0, 1]$ (this assumption can be avoided in more advanced texts) and densities $\delta_1, ..., \delta_d : \mathbb{R} \to [0, \infty)$ with respect to the Lebesgue measure. The assumption also grants the existence of inverses $\Delta_i^{-1} : [0, 1] \to \mathbb{R}$. We can denote the random vector consisting of these variables as $\boldsymbol{\xi} : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}^d)$ with a multivariate distribution function $\Delta : \mathbb{R}^d \to [0, 1]$ and density $\delta : \mathbb{R}^d \to [0, \infty)$. It is a well known observation, that plugging a random variable into its own distribution function results in a uniform distribution on the interval $[0, 1]$, i.e. $\Delta_i(\boldsymbol{\xi_i}) \sim \mathcal{U}[0, 1]$ for every $i = 1, ..., d$. We can proceed to define the notion of a copula.

**Definition 5.** *(Ruppert and Matteson (2011) ) We define* ***copula*** *as a multivariate cumulative distribution function, whose marginal distributions are all* $\mathcal{U}[0,1]$. *Under the assumption mentioned above, we can define the copula of a random vector* $\boldsymbol{\xi}$ *as a function* $\mathcal{C} : [0,1]^d \to [0,1]$, *such that for each* $u_1, ..., u_d \in [0,1]$ *we have*

$$\mathcal{C}(u_1, ..., u_d) = \mathbb{P}(\Delta_1(\boldsymbol{\xi_1}) \leq u_1, ..., \Delta_d(\boldsymbol{\xi_d}) \leq u_d),$$

*or equivalently, under the assumption which grants the inverse of distribution functions* $\Delta_i$,

$$\mathcal{C}(u_1, ..., u_d) = \mathbb{P}(\boldsymbol{\xi_1} \leq \Delta_1^{-1}(u_1), ..., \boldsymbol{\xi_d} \leq \Delta_d^{-1}(u_d)),$$

As we can see from the definition, copula indeed contains no information about the marginal distributions. This is a good thing, because by fixing a copula, we can create joint distributions with the desired dependence structure only by specifying their marginals. But how can we assemble these two parts together? The answer lies within the famous Sklar theorem (see Ruppert and Matteson (2011)), which combines the copula with marginal distributions. The joint distribution function $\Delta$ of the random vector $\boldsymbol{\xi}$ for each $z_1, ..., z_d \in \mathbb{R}$ can be written as

$$\Delta(z_1, ..., z_d) = \mathcal{C}(\Delta_1(z_1), ..., \Delta_d(z_d)). \tag{2.4}$$

Since the copula $\mathcal{C}$ is essentially a distribution function with nice properties, we would also like to work with its density (with respect to the Lebesgue measure), which can be computed for each $u_1, ..., u_d \in [0,1]$ as the derivative

$$c(u_1, ..., u_d) = \frac{\partial^d}{\partial u_1 \cdots \partial u_d} \mathcal{C}(u_1, ..., u_d).$$

The whole multivariate density $\delta$ of the vector $\boldsymbol{\xi}$ for each $z_1, ..., z_d \in \mathbb{R}$ can then be obtained by differentiating the formula (2.4):

$$\delta(z_1, ..., z_d) = c(\Delta_1(z_1), ..., \Delta_d(z_d)) \cdot \delta_1(z_1) \cdots \delta_d(z_d). \tag{2.5}$$

*Example.* (Ruppert and Matteson (2011)) Now that we have covered the basics, let us show examples of some of the important copulas (or parametric families of copulas).

- The **independence copula**, which is the copula of $d$ independent random variables.

$$\mathcal{C}^{ind}(u_1, ..., u_d) = u_1 \cdots u_d.$$

- The **co-monotonicity copula**, which is the copula measuring perfect positive dependence

$$\mathcal{C}^M(u_1, ..., u_d) = \min(u_1, ..., u_d).$$

- The **bivariate Gumbel copula** for a parameter $\theta \in [1, \infty)$, which is a copula family modelling upper tail dependence.

$$\mathcal{C}^{Gu}(\theta, u_1, u_2) = e^{-((-\log u_1)^\theta + (-\log u_2)^\theta)^{\frac{1}{\theta}}}.$$

  For $\theta = 1$ it is the independence copula and for $\theta \to \infty$ it is the co-monotonicity copula.

- The **Gaussian copula** for a correlation matrix $\Sigma \in [0, 1]^{d \times d}$, which is the copula of a multivariate normal distribution with the correlation matrix $\Sigma$

$$\mathcal{C}_\Sigma^{Gauss}(u_1, ..., u_d) = \Phi_\Sigma(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)),$$

  where $\Phi_\Sigma$ is a distribution function of $N_d(0, \Sigma)$ and $\Phi$ is a distribution function of $N(0, I_d)$

### 2.6.2 Endogenous randomness modelling

The theoretical basics are covered, now we return back to our problem setting. For each feasible decision $x \in \mathcal{X}_0$ we have the random element $\boldsymbol{\xi}(x) = (\boldsymbol{\xi_1}(x), ..., \boldsymbol{\xi_d}(x))^T$ in our programming model and we assume continuous and strictly increasing distribution functions. Denoted by $\Delta(x, \cdot)$ is the joint distribution function and $\delta(x, \cdot)$ the joint density with respect to the Lebesgue measure. Moreover, $\Delta_j(x, \cdot)$ is the marginal distribution function of $\boldsymbol{\xi_j}(x)$ and $\delta_j(x, \cdot)$ its density. As for the dependence structure, there is a decision dependent copula associated with $\boldsymbol{\xi}(x)$, denoted $\mathcal{C}(x, \cdot)$ with a density $c(x, \cdot)$. This is a very general case, let us break it down to two special cases.

#### Fixed copula

The first considered case of endogenous randomness is, that the decisions influence only the marginal distributions, but preserve their dependence structure. This means that we have a fixed copula $\mathcal{C}$ with density $c$ for each decision $x \in \mathcal{X}_0$ and decision dependent distribution functions. Now we can mend them together and compute the density of $\boldsymbol{\xi}(x)$ using formulas 2.4 and 2.5. For an arbitrary feasible decision $x \in \mathcal{X}_0$ and a realization $\xi = (\xi_1, ..., \xi_d)^T \in \Xi(x)$, we obtain the value of the decision dependent joint density $\delta(x, \xi)$ as

$$\delta(x, \xi) = c\Big(\Delta_1(x, \xi_1), ..., \Delta_d(x, \xi_d)\Big) \cdot \delta_1(x, \xi_1) \cdot \cdots \delta_d(x, \xi_d).$$

This object uniquely specifies the distribution $\mathcal{P}(x)$ for each feasible decision. As for the shape of dependence of the marginal distributions, we can use an arbitrary method from the previous sections. We can finally formulate the whole stochastic optimization problem with a fixed feasibility set and fixed copula as

$$\min_{x\in\mathcal{X}_0}\int_{\mathbb{R}^d} f(x,\xi)\cdot c\Big(\Delta_1(x,\xi_1),...,\Delta_d(x,\xi_d)\Big)\cdot\delta_1(x,\xi_1)\cdot\cdots\delta_d(x,\xi_d)\,d\xi.$$

*Example.* A useful example of this type of decision dependence are strictly increasing transformations. Let us denote by $\boldsymbol{\xi}=(\boldsymbol{\xi_1},...,\boldsymbol{\xi_d})^T$ some basic random vector with a copula $\mathcal{C}$. As we can see in Ruppert and Matteson (2011), under the assumptions of this section, strictly increasing transformations preserve copulas. Let us consider strictly increasing decision dependent transformations $T_1(x,\cdot),...,T_d(x,\cdot):\mathbb{R}\to\mathbb{R}$. We can set the dependence of distributions in the following way:

$$\boldsymbol{\xi}(x)=(\boldsymbol{\xi_1}(x),...,\boldsymbol{\xi_d}(x))^T\overset{\text{a.s.}}{=}(T_1(x,\boldsymbol{\xi}_1),...,T_d(x,\boldsymbol{\xi}_d))^T.$$

Then, for different feasible decisions does $\boldsymbol{\xi}(x)$ generally have different marginals, but the dependence structure defined by the copula stays the same.

### 2.6.3 Fixed marginals

Same as in Section 2.5, we can consider the previous case in reverse. Let us now assume, that each feasible decision preserves the marginal distributions but alters the copula. In mathematical terms, we have common marginal distribution functions $\Delta_1,...,\Delta_d$ and densities $\delta_1,...,\delta_d$. For each $x\in\mathcal{X}_0$ does the element $\boldsymbol{\xi}(x)$ have an associated copula $\mathcal{C}(x,\cdot)$ and copula density $c(x,\cdot)$, joint distribution function $\Delta(x,\cdot)$ and joint density $\delta(x,\cdot)$. Utilizing theses facts and formulas (2.4) and (2.5), we arrive at the joint density of $\boldsymbol{\xi}(x)$ with respect to the Lebesgue measure, whose value for a realization $\xi=(\xi_1,...,\xi_d)^T\in\Xi(x)$ is

$$\delta(x,\xi)=c\Big(x,(\Delta_1(\xi_1),...,\Delta_d(\xi_d))\Big)\cdot\delta_1(\xi_1)\cdot\cdots\delta_d(\xi_d),$$

which uniquely specifies the decision dependent distribution $\mathcal{P}(x)$. The whole optimization problem for a fixed feasibility set and fixed marginal distributions is written as

$$\min_{x\in\mathcal{X}_0}\int_{\mathbb{R}^d} f(x,\xi)\cdot c\Big(x,(\Delta_1(\xi_1),...,\Delta_d(\xi_d))\Big)\cdot\delta_1(\xi_1)\cdot\cdots\delta_d(\xi_d)\,d\xi.$$

The question now is, how to model the dependence of copulas on decisions. The most natural way is to keep the copulas inside some parametric family. One of

them was mentioned in the example on page 25, the bivariate Gumbel copula, which has a single parameter $\theta \in [1, \infty)$. The Gumbel copula is reserved for positive dependencies between two variables, capable of modelling tail risk, and the parameter value spans from absolute independence to perfect positive dependence. We could create a parametric function $\theta : \mathcal{X}_0 \to [1, \infty)$ and set the dependence as $\mathcal{C}(x, \cdot) = \mathcal{C}^{Gu}(\theta(x), \cdot)$, so that the decisions influence the strength of positive dependence between two random elements, but do not alter their marginal distributions.

## 2.7 Set-valued dependence

Let us still consider the desired subset $\mathcal{P}_0$ of all distributions on $(\mathbb{R}^d, \mathcal{B}^d)$. Up to this moment, we have been operating under the assumption that for each feasible decision $x \in \mathcal{X}_0$ there is a uniquely assigned probability distribution $\mathcal{P}(x)$ of the random element. There are several types of situations, where this assumption does not hold. The existence of endogenous randomness does not necessarily implicate, that the decisions fully determine the distribution. There are cases, where they only affect the distribution, but there is still some external influence bringing additional uncertainty to the model, which the decision maker can not control. However, this is not the whole story, because it can also happen that the set is assigned by the decision maker themselves, in order to account for possible uncertainty in the assigned distribution (the specific shape of the distribution might depend on some parameters, that are hard to estimate or account for). Our basic expected value formulation (2.3) does not make sense under the new assumption. As we can see in Jonsbråten et al. (1998), the proposed generalized shape, assuming a fixed feasibility set, is

$$\begin{aligned} \min_{x \in \mathcal{X}_0} \; & \int_{\mathbb{R}^d} f(x, \xi)\, d\mathcal{P}(\xi) \\ s.t. \; & (\mathcal{P}, x) \in \mathcal{K}. \end{aligned}$$

The set $\mathcal{K} \subseteq \mathcal{P}_0 \times \mathcal{X}_0$ is linking the distributions to feasible decisions. The linking constraints can generally be very complex, so we have to limit ourselves to manageable cases. Note that under our previous assumption of uniqueness, the set $\mathcal{K}$ would be of the shape $\mathcal{K} = \{(\mathcal{P}(x), x) : x \in \mathcal{X}_0\}$, i.e. the graph of the dependence mapping $x \to \mathcal{P}(x)$. To proceed formulating new cases, we will now assume that for every feasible decision $x \in \mathcal{X}_0$ there is an assigned set $\mathcal{U}(x) \subseteq \mathcal{P}_0$ of possible distributions of the random element $\boldsymbol{\xi}(x)$. Under this assumption we can consider the linkage set to be $\mathcal{K} = \{(\mathcal{P}, x) : x \in \mathcal{X}_0,\ \mathcal{P} \in \mathcal{U}(x)\}$. We will focus on two methods for solving such problems. The first method is to use robust reformulations, specifically the worst-case approach. The second method is to aggregate the possible distributions into one distribution $\mathcal{P}(x)$ and proceed to solve it the classical way. Let us present some of the new cases, useful under our new assumption. Note that, same as before, this is not supposed to be an exhaustive list of options, merely a demonstration of a few useful ones.

### 2.7.1 Fixed and finite cardinality of $\mathcal{U}(x)$

We will consider that the set $\mathcal{U}(x)$ is finite for all $x \in \mathcal{X}_0$ with a fixed cardinality $k \in \mathbb{N}$. In other words, we assume that for each $x \in \mathcal{X}_0$ it holds that

$$\mathcal{U}(x) = \{\mathcal{P}_1(x), ..., \mathcal{P}_k(x)\}.$$

The first option is to use the robust formulation

$$\min_{x \in \mathcal{X}_0} \max_{\mathcal{P} \in \mathcal{U}(x)} \mathbb{E}_{\mathcal{P}}[f(x, \boldsymbol{\xi}(x))] = \min_{x \in \mathcal{X}_0} \max_{\mathcal{P} \in \mathcal{U}(x)} \int_{\mathbb{R}^d} f(x, \xi) d\mathcal{P}(\xi),$$

which amounts to minimizing the cost of the worst-case scenario. A possible downside of this approach is that it is often too conservative. It is not always wise to consider the worst possible case and it highly depends on the nature of the problem and the risk aversion of the decision maker.

A more refined approach would be to aggregate the distributions from $\mathcal{U}(x)$ into one. If the decision maker does not have any additional information about the mechanism, which selects the true candidate distribution, the most natural choice is to consider the simple average of the distributions. This means, that for every feasible decision, the endogenous distribution will be chosen as

$$\mathcal{P}(x) = \frac{1}{k} \sum_{i=1}^{k} \mathcal{P}_i(x).$$

This phenomenon is called a mixture distribution. The formula means that for every measurable set $B \in \mathcal{B}^d$ it holds that $\mathcal{P}(x)(B) = \frac{1}{k} \sum_{i=1}^{k} \mathcal{P}_i(x)(B)$. We can think of it as throwing a fair dice which tells us which distribution to use. The upside of this approach is that its easy to work with and does not require any additional information. The downside is that the mixture distribution likely does not even belong to the set $\mathcal{U}(x)$.

Cases like this often arise when the decision maker can influence the random element, but there is still some other exogenous random influence, which the decision maker can not control. Let us now assume the existence of an exogenous random element $\boldsymbol{\eta} : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ with a finite set of scenarios $\eta^1, ..., \eta^k \in E$, whose realization has a direct influence on the random element $\boldsymbol{\xi}(x)$. Each feasible decision $x \in \mathcal{X}_0$ then generates the set of candidate distributions, all of them conditional on a realization of the external random element $\boldsymbol{\eta}$. For each $x \in \mathcal{X}_0$ we consider

$$\mathcal{U}(x) = \{\mathcal{P}(x|\eta^1), ..., \mathcal{P}(x|\eta^k)\},$$

where $\mathcal{P}(x|\eta^i)$ denotes the conditional distribution of $\boldsymbol{\xi}(x)$ under the condition $\boldsymbol{\eta} = \eta^i$. Ideally, the decision maker should be aware of this external influence and if there is no further information available about it, they can either use the worst-case formulation or simply aggregate the candidate distributions by a simple

average. However, when the decision maker knows the scenario probabilities $p^i = \mathbb{P}(\boldsymbol{\eta} = \eta^i)$, either true or estimated, they can naturally use them as weights in the mixture distribution. The sum of the weights is one, so it is a correctly defined probability distribution. The final assigned distribution is a weighted average of the partial distributions and defined as

$$\mathcal{P}(x) = \sum_{i=1}^{k} p^i \cdot \mathcal{P}(x|\eta^i).$$

The full model formulation under this setting is the classical one and we can write it as

$$\begin{aligned}
&\min_{x \in \mathcal{X}_0} \ \mathbb{E}_{\mathcal{P}(x)}[f(x, \boldsymbol{\xi}(x))] \\
=&\min_{x \in \mathcal{X}_0} \ \int_{\mathbb{R}^d} f(x, \xi) d\mathcal{P}(x)(\xi) \\
=&\min_{x \in \mathcal{X}_0} \ \sum_{i=1}^{k} p^i \cdot \int_{\mathbb{R}^d} f(x, \xi) d\mathcal{P}(x|\eta^i)(\xi) \\
=&\min_{x \in \mathcal{X}_0} \ \sum_{i=1}^{k} p^i \cdot \mathbb{E}_{\mathcal{P}(x|\eta^i)}[f(x, \boldsymbol{\xi}(x))].
\end{aligned}$$

Note that it is also theoretically possible to work with distributions of the external element $\boldsymbol{\eta}$ which have an infinite or even uncountable amount of scenarios. Instead of a finite vector of probabilities, we use a sequence of them (for the countable case) or a density function (for the uncountable case). This way we can create a mixture distribution of infinitely many distributions.

*Example.* (Motivation) Imagine a vendor who wants to open an ice cream stand for an outdoor weekend festival. The vendor wants to maximize profit by choosing the optimal price for a scoop. Assume two positive-valued functions $a, b$. We will consider the ice cream demand to be a source of endogenous randomness, whose distribution is affected by the scoop price $p$. However, the price is not the only influence on the demand, there is also an important external factor, the weather. The vendor thinks about the weather as a categorical variable $\boldsymbol{\eta}$ with three possible states: $\eta^1 = rainy$, $\eta^2 = windy$ and $\eta^3 = sunny$. Since the event is outdoors, ice cream is more demanded in warmer weather. The vendor knows, that for a price $p$, the demand has a uniform distribution, which

- in rainy weather is $\mathcal{P}(x|\eta^1) = \mathcal{U}[\frac{1}{2}a(p), \frac{1}{2}b(p)]$,
- in windy weather is $\mathcal{P}(x|\eta^2) = \mathcal{U}[a(p), b(p)]$,
- in sunny weather is $\mathcal{P}(x|\eta^3) = \mathcal{U}[\frac{3}{2}a(p), \frac{3}{2}b(p)]$.

Luckily, the vendor has studied the weather forecast and estimated the probability of a rainy weather as $p^1 = \frac{1}{5}$, windy weather as $p^2 = \frac{3}{10}$ and sunny weather

occurring as $p^3 = \frac{1}{2}$. Based on this information, the best course of action is to optimize the model with endogenous random demand while using the distribution $\mathcal{P}(x) = \frac{1}{3} \cdot \mathcal{P}(x|\eta^1) + \frac{3}{10} \cdot \mathcal{P}(x|\eta^2) + \frac{1}{2} \cdot \mathcal{P}(x|\eta^3)$ for each price in question.

### 2.7.2 Ambiguity sets

Earlier on we were mentioning a potential problem with the robust worst-case approach, that it can be too conservative to hedge against the worst of many options. However, it can be a very reasonable method when the set of distributions is rather small, but not in terms of cardinality, but in terms of how different the distributions inside the set are. In this context are the sets $\mathcal{U}(x)$ called **ambiguity sets** and they are commonly used in robust formulations of stochastic optimization problems. Ambiguity sets are usually constructed as neighborhoods of some reference distribution. We have to realize that the true probability distributions in question are rarely known by the decision maker, therefore they have to be estimated. This always yields some sort of error and uncertainty about the estimated quantities. But what if only a small deviation of the estimated distribution from the true one gives us tremendously different results? This is why we construct the neighborhood around the reference distribution and hedge against the worst-case scenario, in order to mitigate the risk of highly suboptimal decisions stemming from an incorrectly specified model. We could also understand ambiguity sets as sets of distributions, which the decision maker deems plausible for a given decision (hence the word "ambiguity"). The thing we have not yet specified is, what do we mean by the neighborhood of a distribution. To answer that, we need to introduce some notion of distance between probability measures (which is not necessarily a metric). This is where a wide variety of ambiguity sets comes into form. Following the papers Luo and Mehrotra (2020) and Basciftci et al. (2021), we present some useful examples of ambiguity sets, all of which were demonstrated on practical examples in those papers. Let us say, that for each feasible decision $x \in \mathcal{X}_0$ there is a reference distribution $\mathcal{P}(x) \in \mathcal{P}_0$ of the random element and we want to incorporate the uncertainty regarding its exact shape. Note that we will provide only the raw shapes of the ambiguity sets, not their reformulations for specific purposes.

#### Ambiguity sets induced by the Wasserstein metric

This type of ambiguity sets was taken from the paper Luo and Mehrotra (2020). Let us first define the Wasserstein metric, which is a measure of probabilistic distance. It is going to be a function of two arguments, both of which are probability distributions from $\mathcal{P}_0$. Denote by $\mathcal{P}(\mathbb{R}^d, \mathcal{B}^d)$ the set of distributions on this space. The metric is denoted by $\mathcal{W} : \mathcal{P}_0 \times \mathcal{P}_0 \to \mathbb{R}$. First we need to define a function

$$\mathcal{S}(\mathcal{P}_1, \mathcal{P}_2) =$$
$$\Big\{\mathcal{P} \in \mathcal{P}(\mathbb{R}^{d+d}, \mathcal{B}^{d+d}) \;\Big|\; \forall B \in \mathcal{B}^d : \mathcal{P}(B \times \mathbb{R}^d) = \mathcal{P}_1(B),\ \mathcal{P}(\mathbb{R}^d \times B) = \mathcal{P}_2(B)\Big\},$$

where $\mathcal{B}^{d+d} = \mathcal{B}^d \otimes \mathcal{B}^d$. This essentially means that for two distributions $\mathcal{P}_1, \mathcal{P}_2$ we obtain a set of all joint probability distributions whose marginals are $\mathcal{P}_1, \mathcal{P}_2$. Then, for an arbitrary norm $||\cdot||$ on $\mathbb{R}^d$, we can define the Wasserstein metric as

$$\mathcal{W}(\mathcal{P}_1, \mathcal{P}_2) = \inf_{\mathcal{P} \in \mathcal{S}(\mathcal{P}_1, \mathcal{P}_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} ||\xi_1 - \xi_2||\, d\mathcal{P}(\xi_1 \times \xi_2).$$

For $x \in \mathcal{X}_0$ we can finally define the ambiguity set as

$$\mathcal{U}^W(x) = \Big\{\mathcal{P} \in \mathcal{P}_0 \;\Big|\; \mathcal{W}(\mathcal{P}, \mathcal{P}(x)) \leq \epsilon(x)\Big\},$$

which means that acceptable distributions for a feasible decision $x$ lie in the $\epsilon(x)$ neighborhood of the reference distribution $\mathcal{P}(x)$. In a special case we could consider the diameter $\epsilon$ to be independent of the decisions.

**Ambiguity sets induced by $\phi$-divergence**

This type of ambiguity set was taken from the paper Luo and Mehrotra (2020). The principle will be the same, only using a different measure of probabilistic distance. Let $\phi : (0, \infty) \to \mathbb{R}$ be a convex function such that $\phi(1) = 0$ and $\mathcal{P}_1, \mathcal{P}_2 \in \mathcal{P}_0$ two distributions such that $\mathcal{P}_1 \ll \mathcal{P}_2$. Under the assumptions above, we can define the $\phi$-divergence as

$$\mathcal{D}_\phi(\mathcal{P}_1 || \mathcal{P}_2) = \mathbb{E}_{\mathcal{P}_2}\Big[\phi\Big(\frac{d\mathcal{P}_1}{d\mathcal{P}_2}\Big)\Big],$$

where $\frac{d\mathcal{P}_1}{d\mathcal{P}_2}$ denotes the Radon-Nikodym derivative. Note that $\phi$-divergence is not a metric, since it is not symmetrical. A famous special case for $\phi(t) = t \cdot \log(t)$ is the Kullback-Leibler divergence, also called relative entropy. The corresponding ambiguity set can be defined as

$$\mathcal{U}^\phi(x) = \Big\{\mathcal{P} \in \mathcal{P}_0 \;\Big|\; \mathcal{D}_\phi(\mathcal{P} || \mathcal{P}(x)) \leq \epsilon(x)\Big\},$$

where we consider the neighborhood of the reference distribution with respect to the new measure of distance.

**Scenario based ambiguity sets**

This ambiguity set was inspired by the paper Basciftci et al. (2021) about a distributionally robust facility location problem. We will provide a highly simplified version. This one was chosen because it assumes a fixed and finite number of scenarios of the random element, i.e. $\cup_{x\in\mathcal{X}_0}\Xi(x) = \{\xi_1, ..., \xi_S\}$. For each feasible decision is the reference distribution $\mathcal{P}(x)$ determined by the decision dependent scenario probabilities $p^1(x), ..., p^S(x)$. We want to incorporate uncertainty about the exactness of the probabilities and for that we can design a simple ambiguity set of the form

$$\mathcal{U}^S(x) = \Big\{\mathcal{P} \in \mathcal{P}_0 \text{ with probabilities } \{p^s\}_{s=1}^S \;\Big|\; |p^s - p^s(x)| \leq \epsilon(x)\ \forall s = 1, ..., S\Big\}.$$

Other types of ambiguity sets may rely on the Kolmogorov-Smirnov distance, moment based inequalities and many others. All the possibilities require suitable reformulations, which can be very complex, but the common approach to solving problems of this type is called the **distributionally robust** approach. We can reformulate the three problems with fixed feasibility sets as

$$\begin{aligned}
&\min_{x\in\mathcal{X}_0} \max_{\mathcal{P}\in\mathcal{U}^W(x)} \int_{\mathbb{R}^d} f(x,\xi)\, d\mathcal{P}(\xi),\\
&\min_{x\in\mathcal{X}_0} \max_{\mathcal{P}\in\mathcal{U}^\phi(x)} \int_{\mathbb{R}^d} f(x,\xi)\, d\mathcal{P}(\xi),\\
&\min_{x\in\mathcal{X}_0} \max_{\mathcal{P}\in\mathcal{U}^S(x)} \int_{\mathbb{R}^d} f(x,\xi)\, d\mathcal{P}(\xi),
\end{aligned}$$

under their respective settings and assumptions. If the ambiguity set is reasonably small (mainly the parameters $\epsilon(x)$), the solution hedges well against inaccuracies in the estimated distribution, but at the same time, it is not overly conservative.

### 2.7.3 Contamination

In the previous section we considered the sets of decision dependent distributions to be constructed as neighborhoods of some reference distribution. Another very common shape of a decision dependent set is a line segment between two distributions. To elaborate more, we often face the danger of misspecification of the underlying distribution, since it is usually estimated from data and thus vulnerable to errors. The contamination approach, as a method of stress testing, provides a framework for assessing the solutions stability. It is crucial to study how small changes in the distribution affect the optimal solution and its objective value. If the problem is highly sensitive to the specific shape of the distribution, any misspecification might have detrimental effects on the validity of the optimal solution. As we will see in the following text, contamination can transform the stability examination into a problem of one parameter $\lambda$.

First, let us demonstrate the concept on a classical stochastic programs with exogenous randomness. This section is heavily inspired by the paper Dupačová (2006). In the first chapter we considered a random element $\boldsymbol{\xi}$ with a probability distribution $\mathcal{P}$. Now we want to evaluate the sensitivity of the optimal solution with respect to changes in the underlying distribution. We can model that by considering an alternative distribution $\tilde{\mathcal{P}}$. We will contaminate our distribution with the alternative one by using their convex combination, i.e. for every $\lambda \in [0, 1]$ we consider the contaminated distribution

$$\mathcal{P}(\lambda) = (1 - \lambda) \cdot \mathcal{P} + \lambda \cdot \tilde{\mathcal{P}}.$$

These distributions for all values $\lambda \in [0, 1]$ construct the line segment between $\mathcal{P}$ and $\tilde{\mathcal{P}}$. For $\lambda = 0$ we obtain our reference distribution, and raising the parameter slowly contaminates it by the alternative one. Under the assumption that the optimal solution exists for every distribution on this line, we can denote the objective function as

$$F(x, \lambda) = \mathbb{E}_{\mathcal{P}(\lambda)}[f(x, \boldsymbol{\xi})] = (1 - \lambda) \cdot \mathbb{E}_{\mathcal{P}}[f(x, \boldsymbol{\xi})] + \lambda \cdot \mathbb{E}_{\tilde{\mathcal{P}}}[f(x, \boldsymbol{\xi})].$$

Since it is linear in $\lambda$, by differentiating with respect to $\lambda$ we get

$$\frac{\partial}{\partial \lambda} F(x, \lambda) = \mathbb{E}_{\tilde{\mathcal{P}}}[f(x, \boldsymbol{\xi})] - \mathbb{E}_{\mathcal{P}}[f(x, \boldsymbol{\xi})],$$

which is the change in the objective function after moving from $\mathcal{P}$ to $\tilde{\mathcal{P}}$. For the optimal value function $\phi(\lambda) = \min_{x \in \mathcal{X}_0} F(x, \lambda)$ and its **unique** minimizer $\hat{x}$, under some mild conditions there exist the one-sided derivative

$$\phi'(0^+) = \mathbb{E}_{\tilde{\mathcal{P}}}[f(\hat{x}, \boldsymbol{\xi})] - \mathbb{E}_{\mathcal{P}}[f(\hat{x}, \boldsymbol{\xi})] = \frac{\partial}{\partial \lambda} F(\hat{x}, \lambda).$$

This means that the local change of the optimal value after a slight movement in the direction $\tilde{\mathcal{P}} - \mathcal{P}$ corresponds to the change of the objective function after switching the distribution at the minimizer $\hat{x}$. The function $\phi(\lambda)$ is also concave in $\lambda$ which gives us valid bounds for each $\lambda \in [0, 1]$

$$\phi(\lambda) \in \Big[(1 - \lambda) \cdot \phi(0) + \lambda \cdot \phi(1),\ \phi(0) + \lambda \cdot \phi'(0^+)\Big],$$

which we can use to assess the stability of the optimal value, since it gives us an interval in which the optimal value stays under each distribution on the whole line segment. The decision dependent case is much more complicated.

Let $F(x, \mathcal{P}(x)) = \mathbb{E}_{\mathcal{P}(x)}[f(x, \boldsymbol{\xi}(x))]$ be the standard objective function for endogenous randomness and $\hat{x}$ the true or approximate minimizer. Now for every feasible solution $x \in \mathcal{X}_0$ we contaminate the distribution $\mathcal{P}(x)$ by $\tilde{\mathcal{P}}(x)$ and the sensitivity can again be measured by the derivative of $F(\hat{x}, \mathcal{P}(x))$ in the direction $\tilde{\mathcal{P}}(x) - \mathcal{P}(x)$. There is no simple general assertion about the optimal value function anymore, only in special cases, hence there is a need for various algorithmic methods. The goal is again to construct the bounds as before and literature offers a variety of methods. More information on this topic can be found in Kopa and Rusý (2023) for example.

In the context of the whole section, we can consider the set of viable distributions for a decision $x \in \mathcal{X}_0$ to be

$$\mathcal{U}(x) = \{(1 - \lambda) \cdot \mathcal{P}(x) + \lambda \cdot \tilde{\mathcal{P}}(x) \mid \lambda \in [0, 1]\}.$$

## 2.8 Multi-stage programs with endogenous randomness

In Section 1.3.2 we considered $T$ decision stages induced by the decision vectors $x = (x_1, ..., x_T)^T$, where $x_t \in \mathbb{R}^{n_t}$ and the histories were denoted $x_{[t]} = (x_1, ..., x_t)^T$. The causal pathway of the decision process was

$$\text{decide } x_1 \rightsquigarrow \text{observe } \xi_1 \rightsquigarrow \text{decide } x_2 \rightsquigarrow ... \rightsquigarrow \text{observe } \xi_{T-1} \rightsquigarrow \text{decide } x_T.$$

The difference now is, that the random elements depend on the decisions. But that's not the whole story. Classical multi-stage programming is complicated and computationally expensive as it is, but adding endogenous randomness brings another layer of complexity. The simpler case we can consider is assumes, that the decisions do not alter the nature of the decision process and the times at which the random elements are observed. In relation to nonanticipativity, the random element after the $t$-th stage decision depends on the history of decisions up to that stage, meaning that the causal pathway is now

$$\begin{aligned}\text{decide } x_1 \rightsquigarrow \text{observe } \xi_1(x_1) \rightsquigarrow \text{decide } x_2 \rightsquigarrow ... \\ \rightsquigarrow \text{observe } \xi_{T-1}(x_{[T-1]}) \rightsquigarrow \text{decide } x_T.\end{aligned}$$

The distributions can be affected in various ways that we considered in this chapter. If we assume that the scenario tree has the same shape for each distribution, it is usually tractable. The real trouble starts when these assumptions do not hold anymore. There is a lot of practical problems which require more complex

modelling techniques, because they include situations, where a decision might not only change a distribution, but also the stages at which the decision maker obtains new information about some of the random elements. This can result in the decision maker dealing with completely different tree structures for each decision and the need to look for special problem structures. This is highly non-trivial and not really in the scope of this thesis, but it needs to be mentioned. Example of such a case can be found in the paper Goel and Grossmann (2004), which deals with optimization of an offshore gas field. The uncertainly lies, for example, in the sizes of underground oil reservoirs. One of the decisions is whether to apply a certain device, which can find out more about the amount of oil in a potential reservoir. Such a decision is costly, but provides valuable information, and based on the decision, the scenario tree branches in different ways. There are other practical cases like this and they all exhibit high complexity from both the modelling and computational standpoint.

# 3. Newsvendor problem

The newsvendor problem is one of the most famous models in all of stochastic programming. It serves as a primary example of how randomness can be incorporated into an optimization problem. Note that its use is obviously not limited to selling newspapers. The name stems from the fact that newspapers become worthless after their relevance period expires, and the remaining stock has no remaining value. This is foreshadowing the models key assumption, that the product at hand loses all value after the selling period expires. We will generally talk about a vendor selling some product. The standard formulation of this model contains a random demand, which is treated like an external variable independent of the decisions (exogenous). We will try extending this to an endogenous case. The problem can be found in Hrabec et al. (2012) for example. First let us formulate the basic model.

## 3.1 The standard case

Suppose that a vendor is selling some product and their goal is to choose the optimal order quantity. Since the product becomes worthless after the selling period, the vendor should not overstock, in order to avoid lost profit. The decision variable $x \geq 0$ is the ordered quantity of the product. The ordering cost per unit is a finite number $c > 0$ and the selling price is $p > 0$. We assume that $p > c$, so the trade is profitable. The uncertainty in the model is contained in the product demand, which is represented by a random variable $\boldsymbol{\xi} : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ with a distribution $\mathcal{P}$. The assumption is that the demand realization becomes known after ordering the product stock, with no option of a recourse decision. The goal is to maximize profit as a difference between the gains and costs.

$$\max_{x \geq 0} \; p \cdot \mathbb{E}[\min(x, \boldsymbol{\xi})] - c \cdot x, \tag{3.1}$$

or equivalently

$$\max_{x \geq 0} \; (p - c) \cdot x - p \cdot \mathbb{E}[x - \boldsymbol{\xi}]^+,$$

We will stick with the formulation (3.1) in the remainder of this chapter. It is a known fact that for each demand distribution $\mathcal{P}$ this problem has an explicit optimal solution $x^*(\mathcal{P}) = F_{\mathcal{P}}^{-1}(1 - \frac{c}{p})$, which is the $(1 - \frac{c}{p})$ quantile of the demand distribution. This is often called the critical fractile formula.

Another possible approach would be to use worst-case analysis instead of the expected value criterion, or using probability constraints to control the probability of overstocking. The true shape of the demand distribution $\mathcal{P}$ is rarely known by the vendor, so it needs to be estimated. Note that obtaining historical demand

data could result in sampling from a censored distribution, since the observed demands would be bounded from above by the available stock. To set the basis for the rest of this chapter, let us assume that the demand follows a continuous uniform distribution centered around the expected demand $D = \mathbb{E}[\boldsymbol{\xi}]$ with a volatility parameter $\sigma \geq 0$, specifically

$$\mathcal{P} = \mathcal{U}\Big[D - \frac{\sigma}{2}, D + \frac{\sigma}{2}\Big].$$

## 3.2 Newsvendor problem with continuous price selection

In the model above, we considered a fixed selling price p, but what if the vendor would like to treat the price as a decision variable? This is a realistic concern, but it comes with a complication. The demand will always be a random variable, however, it is not sound to assume that its distribution does not depend on the selling price. This sets the stage for a practical use of decision dependent randomness. More information about this setting can be found in the paper Hrabec et al. (2012).

Now the main question is, how to set the type of dependence. There are several ways of approaching this issue, we could pick the selling price from some discrete set, for example $small-medium-large$, and set the demand distribution for each of these states separately. It would be simple from a computational standpoint, however, we want to work with a continuous selection in order to demonstrate the principles from the previous chapter.

The ordering unit price of the product will again be some finite number $c > 0$. We will consider the possible selling prices from a bounded interval $[a, b]$, where $c < a \leq b < \infty$ (the selling price always has to be larger than the ordering one). Now we have to specify the demand distribution. For each price $p \in [a, b]$ is the demand a random variable $\boldsymbol{\xi}(p) : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ with a distribution $\mathcal{P}(p)$. We will assume that for each $p \in [a, b]$ the demand has a continuous uniform distribution over some bounded interval, where the bounds are decision dependent. For each feasible price we will denote the expected demand as $D(p) = \mathbb{E}[\boldsymbol{\xi}(p)]$. Then we consider a volatility function $\sigma(p) : [a, b] \to [0, \infty]$ (if $\sigma(p) = 0$, we are solving a deterministic program with $D(p)$ as the demand) determining the variance of the demand. For now assume that the demand and variance functions are correctly defined and $D(p) - \frac{\sigma(p)}{2} \geq 0$ for each $p \in [a, b]$. In order to not violate basic economic principles, we will consider the expected demand function to be non-increasing in price. The shape of the volatility function is more debatable, but we will also assume it to be non-increasing, because for a larger price the product attracts a smaller population, which narrows the spread. Finally, we set the decision demand distribution for a price $p \in [a, b]$ as

$$\mathcal{P}(p) = \mathcal{U}\Big[D(p) - \frac{\sigma(p)}{2}, D(p) + \frac{\sigma(p)}{2}\Big],$$

meaning a continuous uniform distribution on the aforementioned interval concentrated around the expected demand. This is the type of dependence studied in Section 2.4 about fixed parametric families. From the moment properties of the uniform distribution, we obtain that the expected value is indeed $D(p)$ and the variance is equal to $\frac{\sigma(p)^2}{12}$ for each feasible price. We will enforce a constraint $x \in \Big[D(p) - \frac{\sigma(p)}{2}, D(p) + \frac{\sigma(p)}{2}\Big]$, which is a valid, because we assume that the vendor knows the bounds of the demand interval and it does not make sense for them to order more than the maximum possible demand, since that would surely result in unsold goods and lost profit. It also does not make sense to order less than the minimum possible demand, since the profit could always be increased with no risk by ordering at least the lower bound $D(p) - \frac{\sigma(p)}{2}$. This argument would not be valid if the vendor could not afford to cover the minimal possible demand, but the problem would lose all meaning in that case. This constraint ensures that the vendor avoids overstocking and understocking and simplifies the objective function. Note that there is a big assumption, that the demand estimates are accurate, which can be difficult in real world. This serves merely as a demonstrative example, but in real life, we could use ambiguity sets to handle the possible error. The new model definition then is

$$\max_{p\in[a,b],\ x\in[D(p)-\frac{\sigma(p)}{2},D(p)+\frac{\sigma(p)}{2}]} p \cdot \mathbb{E}[\min(x, \boldsymbol{\xi}(p))] - c \cdot x.$$

Let us realize that the uniform distribution is absolutely continuous with respect to the one-dimensional Lebesgue measure and for a price $p \in [a, b]$ it has a constant density $\delta(p, z) = \frac{1}{\sigma(p)}$, $z \in \Big[D(p) - \frac{\sigma(p)}{2}, D(p) + \frac{\sigma(p)}{2}\Big]$ (assuming $\sigma(p) > 0$). The problem can then be reformulated as

$$\max_{p\in[a,b],\ x\in\left[D(p)-\frac{\sigma(p)}{2},D(p)+\frac{\sigma(p)}{2}\right]} \frac{p}{\sigma(p)} \cdot \int_{D(p)-\frac{\sigma(p)}{2}}^{D(p)+\frac{\sigma(p)}{2}} \min(x, z)\, dz - c \cdot x, \tag{3.2}$$

and since $x \in \Big[D(p) - \frac{\sigma(p)}{2}, D(p) + \frac{\sigma(p)}{2}\Big]$, we can write

$$
\begin{aligned}
&\int_{D(p)-\frac{\sigma(p)}{2}}^{D(p)+\frac{\sigma(p)}{2}} \min(x,z)\,dz \\
=&\int_{D(p)-\frac{\sigma(p)}{2}}^{x} \min(x,z)\,dz + \int_{x}^{D(p)+\frac{\sigma(p)}{2}} \min(x,z)\,dz \\
=&\int_{D(p)-\frac{\sigma(p)}{2}}^{x} z\,dz + \int_{x}^{D(p)+\frac{\sigma(p)}{2}} x\,dz \\
=&\Big[\frac{z^2}{2}\Big]_{D(p)-\frac{\sigma(p)}{2}}^{x} + x\cdot\Big(D(p)+\frac{\sigma(p)}{2}-x\Big) \\
=&\frac{1}{2}\cdot\Big(x^2-\Big(D(p)-\frac{\sigma(p)}{2}\Big)^2\Big)+x\cdot\Big(D(p)+\frac{\sigma(p)}{2}-x\Big) \\
=&-\frac{1}{2}\cdot x^2+\Big(D(p)+\frac{\sigma(p)}{2}\Big)\cdot x-\frac{1}{2}\cdot\Big(D(p)-\frac{\sigma(p)}{2}\Big)^2 \\
=&-\frac{1}{2}\cdot x^2+\Big(D(p)+\frac{\sigma(p)}{2}\Big)\cdot x-\frac{1}{2}\cdot D(p)^2+\frac{1}{2}\sigma(p)D(p)-\frac{\sigma(p)^2}{8}.
\end{aligned}
$$

After further simplification, we can write the full model (3.2) as

$$
\max_{p\in[a,b],\ x\in\left[D(p)-\frac{\sigma(p)}{2},D(p)+\frac{\sigma(p)}{2}\right]} \alpha(p)\cdot x^2+\beta(p)\cdot x+\gamma(p), \tag{3.3}
$$

where

$$
\begin{aligned}
\alpha(p)&=-\frac{p}{2\cdot\sigma(p)},\\
\beta(p)&=\frac{p}{\sigma(p)}\cdot\Big(D(p)+\frac{\sigma(p)}{2}\Big)-c,\\
\gamma(p)&=\frac{p}{\sigma(p)}\cdot\Big(-\frac{1}{2}\cdot D(p)^2+\frac{1}{2}\cdot\sigma(p)\cdot D(p)-\frac{\sigma(p)^2}{8}\Big).
\end{aligned}
$$

Note that $\alpha(p)<0\ \forall p\in[a,b]$, because $\sigma(p)>0$, which means that for every feasible price is the objective function quadratic and strictly concave in $x$, thus having a unique maximizer. For a fixed feasible price $p$, we can differentiate the objective with respect to $x$ and obtain the equation $2\alpha(p)\cdot x+\beta(p)=0$. We can solve it in the following steps.

$$
\begin{aligned}
2\alpha(p)\cdot x+\beta(p)&=0\\
-\frac{p}{\sigma(p)}\cdot x+\frac{p}{\sigma(p)}\cdot(D(p)+\frac{\sigma(p)}{2})-c&=0\\
\frac{p}{\sigma(p)}\cdot(D(p)+\frac{\sigma(p)}{2})-c&=\frac{p}{\sigma(p)}\cdot x\\
x&=D(p)+\frac{\sigma(p)}{2}-c\cdot\frac{\sigma(p)}{p}\\
x&=D(p)+\left(\frac{1}{2}-\frac{c}{p}\right)\cdot\sigma(p).
\end{aligned}
$$

Since our assumption was, that $c \in (0, p)$ holds for all feasible prices (in order to be able to generate profit), $\frac{1}{2} - \frac{c}{p}$ is a decreasing function in c which takes values in $\left(-\frac{1}{2}, \frac{1}{2}\right)$. This means, that the optimal $x$ always lies inside the feasible interval $\left[D(p) - \frac{\sigma(p)}{2}, D(p) + \frac{\sigma(p)}{2}\right]$ and is therefore optimal and feasible. We can denote the optimal ordering amount for price $p \in [a, b]$ as $x(p) = D(p) + \left(\frac{1}{2} - \frac{c}{p}\right) \cdot \sigma(p)$. From this formula we can see, that for selling prices significantly larger than $c$ is the expected demand very small and, in order to generate profit, the vendor is ordering an amount close to the maximal demand, since the profit from each sold product is very high and the relative cost of unsold goods is not that big. Conversely, for selling prices very close to $c$ does each sold unit generate a small profit and the vendor does not want to risk too much unsold goods, since the relative ordering cost is high, which is why they order an amount close to the minimal demand. The whole optimization problem can then be reformulated using only the price variable as

$$\max_{p \in [a,b]} \ \alpha(p) \cdot x(p)^2 + \beta(p) \cdot x(p) + \gamma(p). \tag{3.4}$$

After obtaining the optimal price $p^*$ from this program, the optimal order quantity is equal to $x^* = x(p^*) = D(p^*) + \left(\frac{1}{2} - \frac{c}{p^*}\right) \cdot \sigma(p^*)$. Recall that in the standard exogenous case, the optimal order quantity was the $(1 - \frac{c}{p})$ quantile of the demand distribution. We can see that this also holds true in this case, i.e. for each feasible price $p$ is the optimal order quantity the $(1 - \frac{c}{p})$ quantile of $\mathcal{P}(p)$. This is why the whole program can be reformulated as a problem of only the price variable and the optimal quantity is automatically determined by the critical fractile formula.

## 3.3 Newsvendor problem with advertisement

Now that the base model is covered, let us add another decision, the option to buy an advertisement for the product. The general idea can be found in the paper Hrabec et al. (2017) for example. There will be a new binary decision variable $y \in \{0, 1\}$ determining whether the vendor buys the advertisement or not. If yes, the advertisement is expected to reach additional $100 \cdot \rho\%$ of the population for some value $\rho > 0$. For simplicity, we treat $\rho$ as a deterministic parameter, which is set beforehand (perhaps even estimated from data). There is a one-time cost $c_A > 0$ associated with the purchase. Since the product will reach additional $100 \cdot \rho\%$ of potential customers under advertisement, we will assume that the expected demand and volatility functions $D, \sigma$ will be raised by the same amount, i.e. by a multiplicative constant $(1 + \rho)$, but only if $y = 1$. For each price $p \in [a, b]$ the new expected demand is $D(p) \cdot (1 + \rho \cdot y)$ and volatility is $\sigma(p) \cdot (1 + \rho \cdot y)$. The new decision dependent distribution of the demand is

$$\mathcal{P}(p,y) \sim \mathcal{U}\Big[(1+\rho\cdot y)\cdot\Big(D(p)-\frac{\sigma(p)}{2}\Big),(1+\rho\cdot y)\cdot\Big(D(p)+\frac{\sigma(p)}{2}\Big)\Big].$$

Since there is only one binary variable in the model, we can simply solve two separate models for $y = 0$ (without advertisement) and $y = 1$ (with advertisement) and then pick the solution with the larger expected profit. The profitability of advertisement will depend on the cost $c_A$, which we will treat as a parameter in the following numerical example. For notation convenience, let us denote $D_A(p) = (1+\rho)\cdot D(p)$ and $\sigma_A(p) = (1+\rho)\cdot\sigma(p)$ as the demand and volatility functions in the advertisement model. The two competing models then are

$$\max_{p\in[a,b],\ x\in\left[D(p)-\frac{\sigma(p)}{2},D(p)+\frac{\sigma(p)}{2}\right]} \frac{p}{\sigma(p)}\cdot\int_{D(p)-\frac{\sigma(p)}{2}}^{D(p)+\frac{\sigma(p)}{2}}\min(x,z)\,dz - c\cdot x$$

and

$$\max_{p\in[a,b],\ x\in\left[D_A(p)-\frac{\sigma_A(p)}{2},D(p)+\frac{\sigma_A(p)}{2}\right]} \frac{p}{\sigma_A(p)}\cdot\int_{D_A(p)-\frac{\sigma_A(p)}{2}}^{D_A(p)+\frac{\sigma_A(p)}{2}}\min(x,z)\,dz - c\cdot x - c_A.$$

Now we need can simplify the new advertisement model. The process is exactly the same as before, only now we need to change the price coefficients $\alpha(p),\beta(p),\gamma(p)$ from the quadratic objective function (3.3). We will create new coefficients $\alpha_A(p),\beta_A(p),\gamma_A(p)$ from the previous ones by plugging in the new demand and volatility functions. Note their only difference, besides the change of functions, is that the advertisement cost $c_A$ needs to be included in the new absolute term $\gamma_A(p)$. The advertisement model is then

$$\max_{p\in[a,b],\ x\in[D_A(p)-\frac{\sigma_A(p)}{2},D_A(p)+\frac{\sigma_A(p)}{2}]} \alpha_A(p)\cdot x^2+\beta_A(p)\cdot x+\gamma_A(p),$$

where

$$\begin{aligned}
\alpha_A(p) &= -\frac{p}{2\cdot\sigma_A(p)},\\
\beta_A(p) &= \frac{p}{\sigma_A(p)}\cdot\Big(D_A(p)+\frac{\sigma_A(p)}{2}\Big)-c,\\
\gamma_A(p) &= \frac{p}{\sigma_A(p)}\cdot\Big(-\frac{1}{2}\cdot D_A(p)^2+\frac{1}{2}\cdot\sigma_A(p)\cdot D_A(p)-\frac{\sigma_A(p)^2}{8}\Big)-c_A.
\end{aligned}$$

For a given feasible price p, the optimal order amount is again denoted $x_A(p) = D_A(p)+\left(\frac{1}{2}-\frac{c}{p}\right)\frac{\sigma_A(p)}{2}$, therefore the final advertisement model is

$$\max_{p\in[a,b]} \alpha_A(p)\cdot x_A(p)^2+\beta_A(p)\cdot x_A(p)+\gamma_A(p).$$

We can compactly write down the whole model including advertisement choice as

$$\begin{aligned}\max_{y\in\{0,1\},\ p\in[a,b]}\ & y\cdot\Big(\alpha_A(p)\cdot x_A(p)^2+\beta_A(p)\cdot x_A(p)+\gamma_A(p)\Big)+\\ & (1-y)\cdot\Big(\alpha(p)\cdot x(p)^2+\beta(p)\cdot x(p)+\gamma(p)\Big)\end{aligned} \tag{3.5}$$

and after obtaining an optimal solution $y^*, p^*$, the optimal ordering amount $x^*$ can be written as

$$y^*\cdot\left(D_A(p^*)+\left(\frac{1}{2}-\frac{c}{p^*}\right)\frac{\sigma_A(p^*)}{2}\right)+(1-y^*)\cdot\left(D(p^*)+\left(\frac{1}{2}-\frac{c}{p^*}\right)\frac{\sigma(p^*)}{2}\right).$$

## 3.4 Numerical example

We will arbitrarily choose the currency to be american dollars ($). Let us assume that the ordering unit price of the product is $c = 10$ and the unit selling price will be chosen from the closed interval $[a, b] = [12, 18]$. This means that the unit profit will vary between 2$ and 8$. As for the demand and volatility functions, they are chosen in the following way.

$$\begin{aligned}D(p)&=1000-100\cdot(p-12),\ p\in[12,18],\\ \sigma(p)&=800-100\cdot(p-12),\ p\in[12,18].\end{aligned} \tag{3.6}$$

From this we can see that the expected demand is linearly decreasing from 1000 to 400 and the volatility is also linearly decreasing from 800 to 200. We chose both functions so they have the same derivative, since we expect the population of potential customers to decrease in the same way in expectation and spread. In Figure 3.1 we can see the demand structure. The $x$-axis covers the feasible unit selling prices and the orange area for each price level shows the demand region. The purple line shows the optimal ordering amounts for each price $p \in [12, 18]$, i.e. $D(p) + \left(\frac{1}{2} - \frac{c}{p}\right)\cdot\sigma(p)$. See that the optimal ordering amounts all fall below the centre of their respective demand intervals, since the critical quantile $1 - \frac{c}{p}$ for our data spans between $\frac{1}{6}$ and $\frac{4}{9}$.

First we will solve the problem without advertisement, i.e. the problem (3.4). We have everything we need to compute the coefficients $\alpha(p), \beta(p), \gamma(p)$ for each feasible price. The results for the optimal solution are summarized in Table 3.1. In Figure 3.2 we can see the dependence of expected profit on unit selling price. The curve is concave and hints us that the solution is a unique maximizer. At

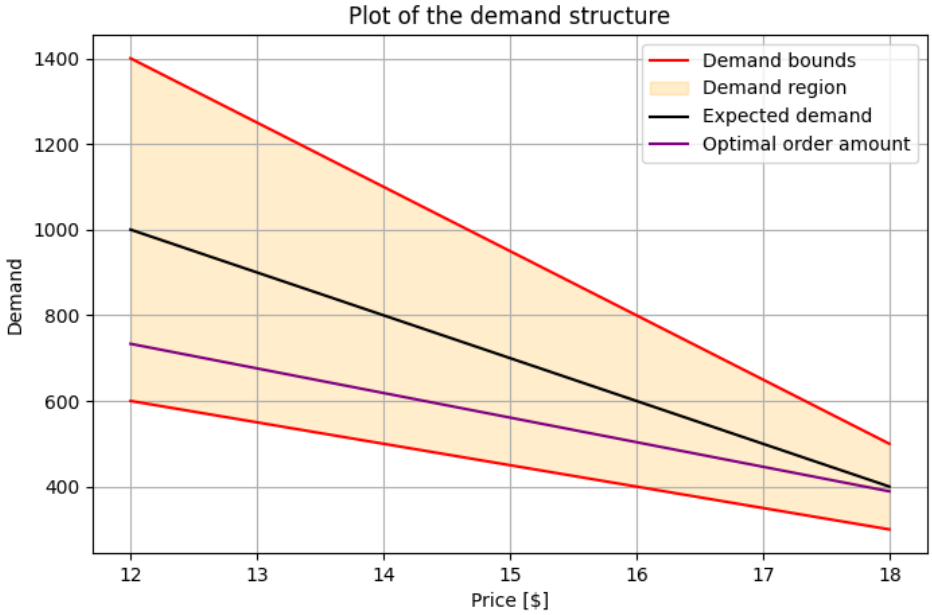


Figure 3.1: Structure of the demand distributions for various price levels

first, raising the price increases the expected profit, but the rate of increase is gradually lowering until it reaches the optimal state. After that the expected profit starts decreasing because of low demand.

| | |
|---|---|
| **Optimal Expected Profit** | 2889.03 $ |
| **Optimal Order Quantity** | 496.68 |
| **Optimal Unit Selling Price** | 16.71 $ |

Table 3.1: Optimal business strategy without advertisement

Now let us insert the binary option of buying an advertisement and solve the problem (3.5). We will set $\rho = 0.2$, so the product reaches additional 20% people, who we assume to have the same chances of buying the product. The demand and volatility functions after purchasing the advertisement are

$$D_A(p) = 1200 - 120 \cdot (p - 12), \ p \in [12, 18],$$
$$\sigma_A(p) = 960 - 120 \cdot (p - 12), \ p \in [12, 18].$$

In Figure 3.3 we can see the comparison of demand intervals with and without advertisement. It is obvious that buying the advertisement can significantly boost the expected profit, but the question is, how much is the vendor willing to pay for it? We discussed earlier that we will consider the cost $c_A$ to be a one-time payment taking place before any of the randomness reveals itself. Instead of choosing an arbitrary advertisement price, we will treat it as a parameter. The expected outcome should be, that if the advertisement is cheap enough, it will be profitable, and after exceeding some turnover price, it will stop being profitable,

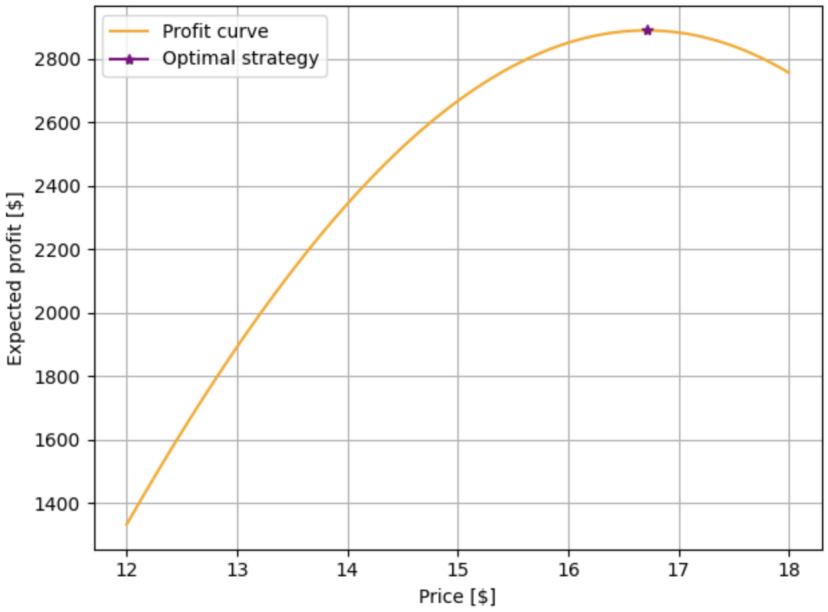


Figure 3.2: Dependence of expected profit on unit price

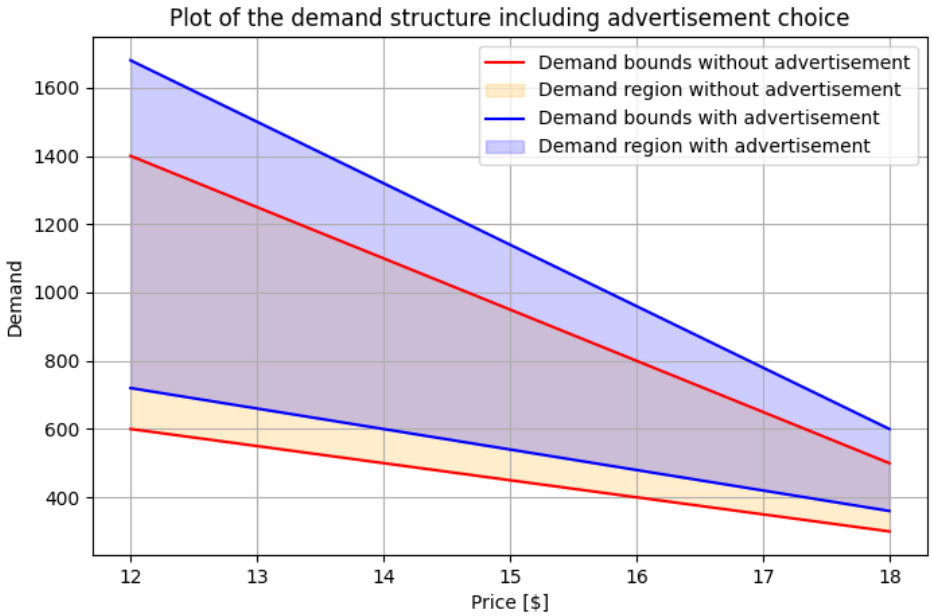


Figure 3.3: Comparison of the demand structure for various price levels

since the costs will exceed the gains. In order to find out the turnover price, we performed a simulation for various values of the advertisement cost, accurate up to one dollar, where we computed the optimal expected profit and benchmarked it against the optimal profit from the model without advertisement.

The result can be seen in Figure 3.4. Since the advertisement cost enters the model as a linear term, the expected profit exhibits a linear decrease. After the price surpasses roughly 577$, it is not profitable to use advertisement anymore, since the profit increase is not enough to cover the costs.

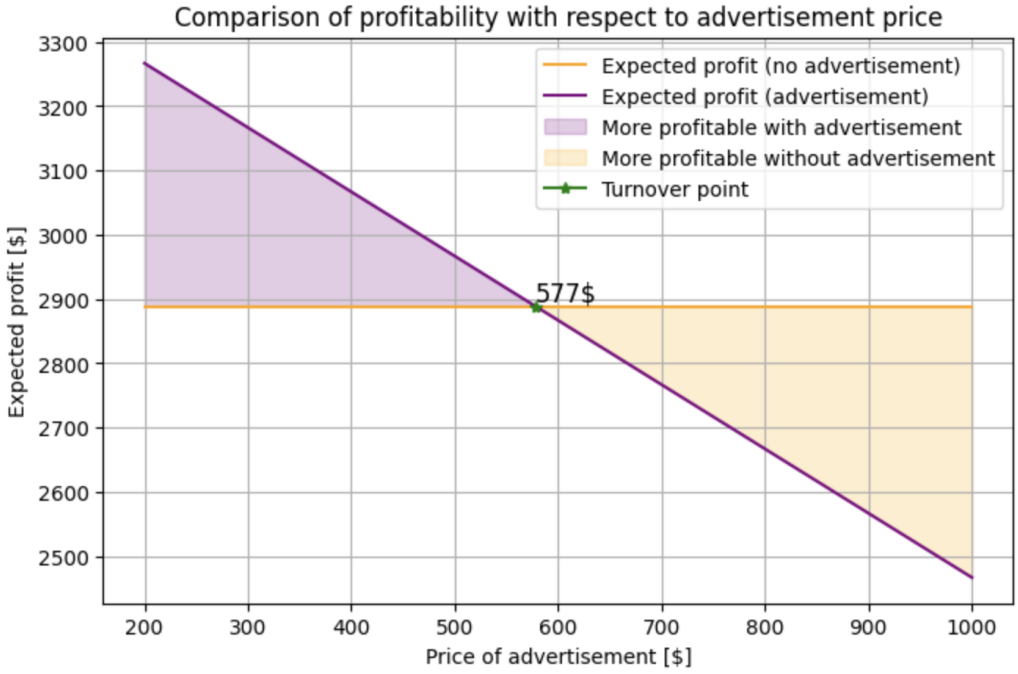


Figure 3.4: Comparison of optimal expected profits with respect to advertisement price

This model was an extension of the the classical newsvendor problem. The goal was to provide a practical formulation for continuous price selection and endogenous randomness. The demand and volatility functions were chosen ad hoc, along with other parameters. The possible downfall of the model simplification is in the assumption that the vendor knows the demand bounds and can act accordingly. In situations like this, it would be a good idea to incorporate uncertainty into the model, possibly by using a distributionally robust formulation with respect to an ambiguity set induced by suitable neighborhoods of the values $D(p)$ and $\sigma(p)$. In the next chapter we will try to extend the situation into more decision stages, giving the vendor an option to react to demand observations by restocking and altering the selling price.

# 4. Three-stage newsvendor problem with recourse

## 4.1 Model formulation

In the previous model we assumed that each unsold product after the given selling period becomes worthless. Because of that, the vendor had to think very carefully about the ordered stock amount. Let us now assume that there are two selling periods. At first, the vendor orders an initial amount of the product and sets a selling price, same as before. However, now the vendor can react to the first stage demand by restocking and altering the selling price, if need be, to prepare for the following selling period. Any remaining product after the whole process loses all value. The ordering price of the product will again be $c > 0$, constant during the whole time horizon. The selling price set by the vendor for the first period will be $p_1$ and for the second period $p_2$, both from the aforementioned positive interval $[a, b]$. We denote $p = (p_1, p_2)^T$ the price process. The two random demands will be represented by a decision dependent random vector $\boldsymbol{\xi}(p) = (\boldsymbol{\xi_1}(p_1), \boldsymbol{\xi_2}(p_2))^T : (\Omega, \mathcal{A}) \to (\mathbb{R}^2, \mathcal{B}^2)$ with marginal distributions $\mathcal{P}_1(p_1)$ and $\mathcal{P}_2(p_2)$ . We will reuse the same notion of expected demand and volatility functions $D$ and $\sigma$ from the previous chapter. Since the final model will belong to the multi-stage framework, we will use a discrete distribution this time. Each random demand will have $S$ possible scenarios for each feasible price, where $S$ is a chosen positive odd integer. The specific values of the demand scenarios will differ for each price level. In the previous chapter it was stated, that for each feasible price p, the distribution of demand was continuous and uniform on the interval $\left[D(p) - \frac{\sigma(p)}{2}, D(p) + \frac{\sigma(p)}{2}\right]$. To obtain the discrete scenarios, we will split these demand intervals into $S$ equidistant states (equidistant because of the original distributions uniformity). The reason for choosing $S$ as an odd number is, that we would like the middle state to be the expected demand $D(p)$. Finally, for some $p \in [a, b]$, both the random demands have a finite equidistant support in the shape

$$\Xi(p) = \left\{ D(p) - \frac{\sigma(p)}{2} + i \cdot \frac{\sigma(p)}{S-1} : \ i \in \{0, ..., S-1\} \right\}.$$

With no additional information about the first stage demand, its distribution will be assumed as uniform on the discrete support set $\Xi(p)$, i.e. $\mathcal{P}_1(p_1) = \mathcal{U}(\Xi(p_1))$. We will represent the first stage scenarios by integers $s_1 = 1, ..., S$ and the first stage probabilities will be denoted $\pi^{s_1} = \frac{1}{S}$. These will be independent of the price level, it is a case of endogenous randomness where the decisions only affect the scenarios, not their probabilities. As for the second stage demand distribution for an updated selling price $p_2$, the scenarios will again come from the set $\Xi(p_2)$, represented by integers $s_2 = 1, ..., S$. Each joint scenario of the demand process can be represented by a pair $s_1, s_2$. Unlike the first stage demand, the conditional

distribution of the second stage demand will generally not be uniform. We will assume that after a specific scenario $s_1$ of the first stage demand occurs, more probability will be concentrated around similar scenarios of the second stage demand. For example, a higher demand in the first stage probably indicates a higher demand in the second stage (by higher we mean relatively to the expected demand for some price level). It is assumed that the prices levels only affect the values of the scenarios, not the dependence structure between them. The transition probabilities between demand scenarios are therefore assumed to be decision independent and they can be stored inside a single matrix $\Pi = (\Pi_{ij})_{i,j=1}^{S}$, where an element $\Pi_{ij} = \pi^{s_i,s_j}$ denotes the conditional probability of transitioning from a first stage scenario $s_1$ to a second stage scenario $s_2$. For example, $\pi^{3,5}$ denotes the conditional probability of observing the fifth lowest second stage demand scenario (for price $p_2$), knowing that the first stage demand was the third lowest scenario (for price $p_1$). This essentially means, that for an arbitrary feasible choice of prices $p_1, p_2$ does the scenario tree of the model have the same topological and probabilistic structure, only with differing scenario values. By $\xi_1^{s_1}(p_1),\ s_1 = 1, ..., S$ we will denote the $s_1$-th scenario of the first stage demand for price $p_1$, analogously for $\xi_2^{s_2}(p_2)$.

We can now proceed to formulate the optimization model. The parameters entering the model are $c, a, b$ and the probabilistic structure $\pi, \Pi$, those are already covered. As for the variables, there will be several. We have three decision stages and two selling periods. We need to create separate decision variables for each branching of the scenario tree, but luckily, our model is defined so that the tree has a fixed shape and transition probabilities between nodes. In order to not violate the nonanticipativity principle, no decision variables can depend on future demand realizations. Let us now summarize the causal pathway of this decision process.

During the fist stage, the vendor chooses $x_1 \geq 0$, the ordered amount, and $p_1 \in [a, b]$, the first selling price. After an arbitrary scenario $\xi_1^{s_1}(p_1)$ of the first stage demand (with an initial probability $\pi^{s_1} = \frac{1}{S}$) becomes known, the vendor makes several second stage recourse decisions. First decision is how much of the available product to sell, denoted by $h_1^{s_1}$. This amount has to be non-negative and it must not be larger than the minimum of the observed demand and available stock. Note that this model does not force the vendor to sell everything they can during the second stage, instead they can keep some portion of the stock for the second selling period, if they deem it a better decision. The amount kept is denoted $s^{s_1}$ and it is the difference of the whole stock and the sold amount. Some models also incorporate a unit holding cost, however, we will not incorporate it here. After all the decisions, the vendor sets a new price $p_2^{s_1} \in [a, b]$ and orders an additional amount $x_2^{s_1} \geq 0$ of the product.

After an arbitrary scenario $\xi_2^{s_2}(p_2^{s_1})$ of the second stage demand (with a conditional probability $\pi^{s_1,s_2}$) becomes known, the last decision is to sell a final amount $h_2^{s_1,s_2} \geq 0$. Obviously, the vendor can not sell more than they have, which is why the amount sold must not be larger than the minimum of the current available stock $x_2^{s_1} + s^{s_1}$ and the second stage demand. All of the model variables and their feasible regions for a scenario pair $s_1, s_2 = 1, ..., S$ are summarized in Table 4.1.

| **1. stage variables** | **Constraints** | **Meaning** |
|---|---|---|
| $\mathrm{p}_1$ | $\in$[a,b] | Selling price for first period |
| $\mathrm{x}_1$ | $\geq 0$ | Order amount for first period |
| **2. stage variables after** $\mathrm{s}_1$ | **Constraints** | **Meaning** |
| $\mathrm{h}_1^{s_1}$ | $\in \left[0,\ \min\left(x_1, \xi_1^{s_1}(p_1)\right)\right]$ | Amount sold in the first period |
| $\mathrm{s}^{s_1}$ | $= \mathrm{x}_1 - h_1^{s_1}$ | Amount kept after the first period |
| $\mathrm{p}_2^{s_1}$ | $\in$[a,b] | Selling price for second period |
| $\mathrm{x}_2^{s_1}$ | $\geq 0$ | Order amount for second period |
| **3. stage variable after** $\mathrm{s}_1, s_2$ | **Constraints** | **Meaning** |
| $\mathrm{h}_2^{s_1,s_2}$ | $\in \left[0,\ \min\left(x_2^{s_1} + s^{s_1}, \xi_2^{s_2}(p_2^{s_1})\right)\right]$ | Amount sold in the second period |

Table 4.1: Three-stage newsvendor model decision variables with feasible regions

The objective of the problem is again to maximize the expected profit. For each scenario pair $s_1, s_2$, the gain is the sum of gains from the two selling periods, i.e. $p_1 \cdot h_1^{s_1} + p_2^{s_1} \cdot h_2^{s_1,s_2}$. The costs are equal to $c \cdot (x_1 + x_2^{s_1})$, the total ordered amount times the unit ordering price (we assumed the ordering price to be the same in both stages, removing the assumption would be simple). We denote all the decision variables using four vectors $p, x, s, h$, which are all non-negative. Utilizing the discrete scenario structure, we formulate the whole problem (4.1) as

$$\begin{aligned}
\max_{p,x,s,h\geq 0} \; & -c \cdot x_1 + \sum_{s_1=1}^{S} \pi^{s_1} \cdot \left[ p_1 \cdot h_1^{s_1} - c \cdot x_2^{s_1} + \sum_{s_2=1}^{S} \pi^{s_1,s_2} \cdot p_2^{s_1} \cdot h_2^{s_1,s_2} \right] \\
& \text{s.t.} \\
& p_1 \in [a,b] \\
& p_2^{s_1} \in [a,b],\ \forall s_1 \in \{1,...,S\} \\
& s^{s_1} = x_1 - h_1^{s_1},\ \forall s_1 \in \{1,...,S\} \\
& h_1^{s_1} \leq x_1,\ \forall s_1 \in \{1,...,S\} \\
& h_1^{s_1} \leq \xi_1^{s_1}(p_1),\ \forall s_1 \in \{1,...,S\} \\
& h_2^{s_1,s_2} \leq s^{s_1} + x_2^{s_1},\ \forall s_1, s_2 \in \{1,...,S\} \\
& h_2^{s_1,s_2} \leq \xi_2^{s_2}(p_2^{s_1})\ \forall s_1, s_2 \in \{1,...,S\}.
\end{aligned} \tag{4.1}$$

After inspecting the shape of the objective function and constraints, we can see that it is a non-convex quadratic program. All the constraints are linear, therefore the feasibility set is convex polyhedral.

## 4.2 Model extension with advertisement

Same as before, we will add the binary possibility of purchasing advertisement. The notation stays the same as before, the one-time cost of advertisement is $c_A > 0$, $y = 1$ corresponds to buying and $y = 0$ not buying. For simplicity, we will assume that the advertisement takes some time to reach the desired effect, therefore it is only effective to buy it in the first decision stage, prior to any demand realization. We will again create the effect parameter $\rho > 0$, such that the demand in the first selling period is raised by a multiplicative factor

$(1+\rho)$. However, we will now assume that the effect can weaken with time. For this we create a decay parameter $\beta \in [0,1]$, such that the advertisement during the second selling period is raising the demand only by a multiplicative factor $(1+\beta\cdot\rho)$. For example, if $\rho = 0.2$ and $\beta = 0.5$, the advertisement raises demand by 20% in the first period and by 10% in the second period. Negative values of the decay parameter would indicate that the advertisement lowers the demand, and values higher than one would indicate that the effect of advertisement gets stronger with time, which is also possible but omitted here. Note that in the context of our definition of the demand distribution, raising the demand by some percentage essentially means, that each scenario is raised by that amount and the probabilistic structure stays the same. Unlike in the previous model (4.1), we have to incorporate the cost of advertisement and change the demand scenarios accordingly. The final model including advertisement can be defined as

$$\begin{aligned}
\max_{p,x,s,h,y\geq 0} & -c_A\cdot y - c\cdot x_1 + \sum_{s_1=1}^{S}\pi^{s_1}\cdot\left[p_1\cdot h_1^{s_1} - c\cdot x_2^{s_1} + \sum_{s_2=1}^{S}\pi^{s_1,s_2}\cdot p_2^{s_1}\cdot h_2^{s_1,s_2}\right] \\
& \text{s.t.} \\
& y \in \{0,1\} \\
& p_1 \in [a,b] \\
& p_2^{s_1} \in [a,b],\ \forall s_1 \in \{1,...,S\} \\
& s^{s_1} = x_1 - h_1^{s_1},\ \forall s_1 \in \{1,...,S\} \\
& h_1^{s_1} \leq x_1,\ \forall s_1 \in \{1,...,S\} \\
& h_1^{s_1} \leq \xi_1^{s_1}(p_1)\cdot(1+\rho\cdot y),\ \forall s_1 \in \{1,...,S\} \\
& h_2^{s_1,s_2} \leq s^{s_1} + x_2^{s_1},\ \forall s_1,s_2 \in \{1,...,S\} \\
& h_2^{s_1,s_2} \leq \xi_2^{s_2}(p_2^{s_1})\cdot(1+\beta\cdot\rho\cdot y)\ \forall s_1,s_2 \in \{1,...,S\}.
\end{aligned}$$

We could also choose a different incorporation of the advertisement effect. For example, the advertisement could alter the scenario probabilities, instead of the values. Moreover, instead of a binary option, we could create a continuous variable, stating how much the vendor invests into advertisement, and model the effect on demand continuously. We chose the binary option for its simplicity and to demonstrate another type of decision dependence modelling. Let us now proceed to a numerical example.

## 4.3 Numerical example

The values $c = 10$, $a = 12$ and $b = 18$ remain the same as before, for the sake comparability. The demand and volatility functions will also be the same, as defined in (3.6). As for the number of demand scenarios, we will choose $S = 9$ to avoid overwhelming numerical complexity. For each price $p \in [12,18]$, the demand scenarios will be referred to by the following code names:

$$
\begin{aligned}
D(p) - \frac{1}{2} \cdot \sigma(p) &= \textit{Lowest} \ , \\
D(p) - \frac{3}{8} \cdot \sigma(p) &= \textit{Very low} \ , \\
D(p) - \frac{1}{4} \cdot \sigma(p) &= \textit{Low} \ , \\
D(p) - \frac{1}{8} \cdot \sigma(p) &= \textit{Below average} \ , \\
D(p) &= \textit{Average} \ , \\
D(p) + \frac{1}{8} \cdot \sigma(p) &= \textit{Above average} \ , \\
D(p) + \frac{1}{4} \cdot \sigma(p) &= \textit{High} \ , \\
D(p) + \frac{3}{8} \cdot \sigma(p) &= \textit{Very high} \ , \\
D(p) + \frac{1}{2} \cdot \sigma(p) &= \textit{Highest} \ .
\end{aligned}
\tag{4.2}
$$

All the first stage probabilities are stored in the vector $\pi$ shown in (4.3). For the transition probabilities we chose the matrix $\Pi$, also shown in (4.3). We assume that after a first stage scenario occurs, it is only possible for the demand to drop or increase by two states or less. Remember that the states themselves depend on the prices, but not the hierarchical structure (4.2) between them. For example, transitioning from a state $s_1 = 1$ to $s_2 = 2$ means, that the first stage demand was the *Lowest* with respect to price level $p_1$, and the second stage demand was *Very low* with respect to price level $p_2^1$. After a first stage demand scenario becomes known, we expect the same scenario in the next stage happening with probability $\frac{4}{10}$, neighboring scenarios with $\frac{2}{10}$ and further neighboring $\frac{1}{10}$. In cases, where there are no more neighbors to either side, on the edges of the matrix, the probabilities aggregate. This approach yields the multidiagonal matrix $\Pi$.

$$
\pi = \begin{pmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{pmatrix}^T ,
$$

$$
\Pi = \begin{pmatrix}
\frac{7}{10} & \frac{2}{10} & \frac{1}{10} & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{3}{10} & \frac{4}{10} & \frac{2}{10} & \frac{1}{10} & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{10} & \frac{2}{10} & \frac{4}{10} & \frac{2}{10} & \frac{1}{10} & 0 & 0 & 0 & 0 \\
0 & \frac{1}{10} & \frac{2}{10} & \frac{4}{10} & \frac{2}{10} & \frac{1}{10} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{10} & \frac{2}{10} & \frac{4}{10} & \frac{2}{10} & \frac{1}{10} & 0 & 0 \\
0 & 0 & 0 & \frac{1}{10} & \frac{2}{10} & \frac{4}{10} & \frac{2}{10} & \frac{1}{10} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{10} & \frac{2}{10} & \frac{4}{10} & \frac{2}{10} & \frac{1}{10} \\
0 & 0 & 0 & 0 & 0 & \frac{1}{10} & \frac{2}{10} & \frac{4}{10} & \frac{3}{10} \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{10} & \frac{2}{10} & \frac{7}{10}
\end{pmatrix} .
\tag{4.3}
$$

The model was fitted using the IPOPT solver, specialized in interior point methods. It uses gradient based optimization to navigate through the feasibility set. Several combinations of feasible starting values were used, all of them yielded the same result. In Figures 4.1 and 4.2 we can see the total optimal expected profits for various combinations of first stage decisions $p_1, x_1$. We can see that with respect to these starting values, the objective function is strictly concave and has a unique maximizer. In the second surface plot we can also see the purple star point corresponding to the optimal first stage strategy, which is described in the following text. Note that these Figures are there just to demonstrate the shape of the objective, there are always orientation problems when projecting into lower dimensions and the optimal solution is not to be read from the plot. They serve as a sort of an extension to Figure 3.2 from the single stage model, merely designed for obtaining some visual context. Note that by "initial values" or "initialization strategy", we mean the first stage decisions, not the starting values entering the solver.

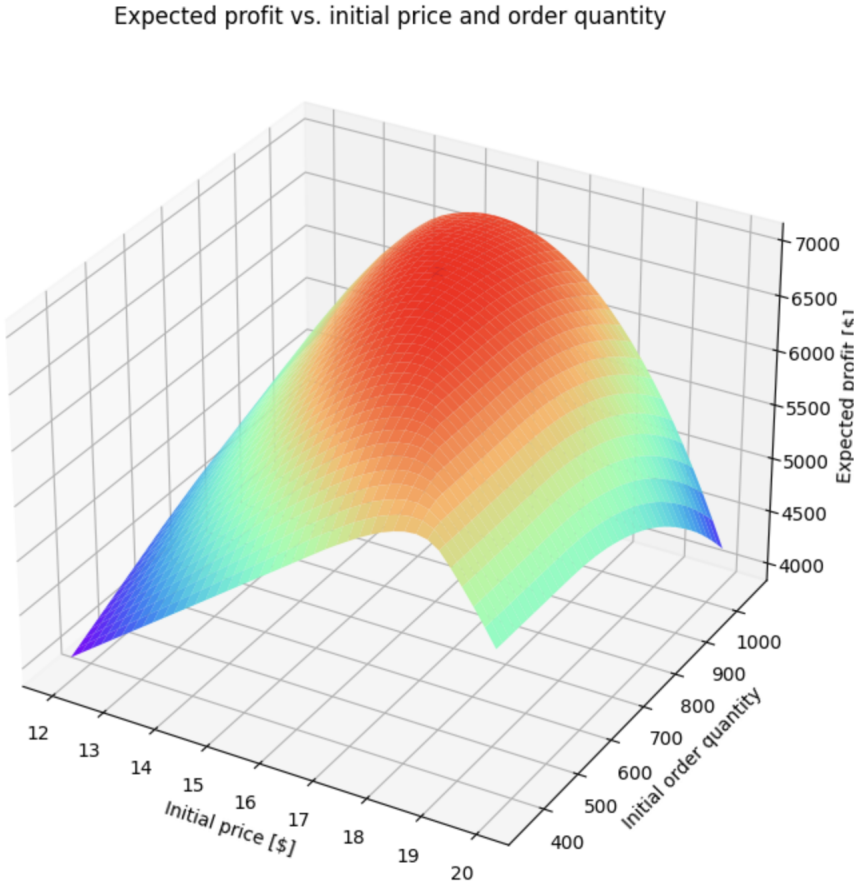


Figure 4.1: Shape of the objective function for different first-stage values

Let us now look at the optimal solution of the model (4.1) without advertisement, summarized in Table 4.2, and try to interpret it. The total optimal expected profit is 6941.17$. Since we used the same price bounds and demand intervals as in the single stage model from the previous chapter (except the previous distribution was continuous, this is discrete), we can loosely compare the optimal profits. Since now there are two periods, we would expect the optimal profit to be roughly

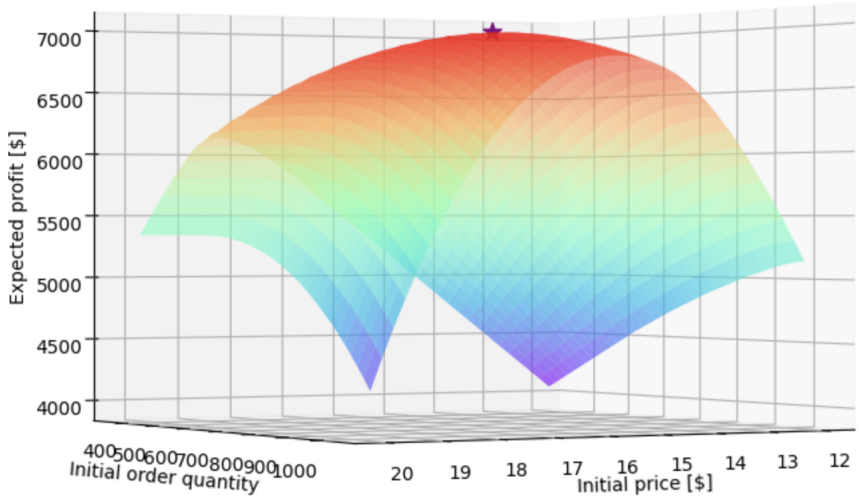


Figure 4.2: The optimal initialization strategy

| $p_1$ | $x_1$ | $\xi_1^{s_1}(p_1)$ | $h_1^{s_1}$ | $s^{s_1}$ | $p_2^{s_1}$ | $x_2^{s_1}$ |
|---|---|---|---|---|---|---|
| | | 395.45 | 395.45 | 390.91 | 16.18 $ | 0.00 |
| | | 444.32 | 444.32 | 342.05 | 17.02 $ | 44.12 |
| | | 493.18 | 493.18 | 293.18 | 16.79 $ | 147.89 |
| | | 542.05 | 542.05 | 244.32 | 16.52 $ | 260.61 |
| 16.09 $ | 786.36 | 590.91 | 590.91 | 195.45 | 16.32 $ | 372.97 |
| | | 639.77 | 639.77 | 146.59 | 16.16$ | 485.10 |
| | | 688.64 | 688.64 | 97.73 | 16.04 $ | 597.06 |
| | | 737.50 | 737.50 | 48.86 | 15.94 $ | 708.92 |
| | | 786.36 | 786.36 | 0.00 | 15.86 $ | 820.69 |

Table 4.2: Optimal business strategy in the three-stage model without advertisement

twice as large as in the single stage model, however, the previous optimal profit was 2889.03$, which is much less than a half. During the single stage business model, the vendor made more conservative decisions about the order amount, since any unsold goods would be discarded after the selling period. In this case, the vendor knows, that unsold goods from the first period can potentionally still be sold during the next one. As we can see, the optimal initial price is quite similar to the previous model, but the order amount in the single stage model is significantly smaller, being the more conservative decision. Table 4.2 has three major parts, separated by double vertical lines. The first part is the first stage strategy, before any demand is observed. The second part tells us how much does the vendor sell and keep during the first selling period. Finally, the third part is about setting the new price and ordering new stock.

The optimal strategy starts by setting the unit price to 16.09$ and ordering 786.36 units. Interestingly, the vendor chose the price in a way that the lowest demand is almost exactly a half of the highest demand, and then ordered the highest possible amount. This means, that while keeping the selling price constant during

| **1. stage scenario $s_1$** | **Most likely 2. stage demand** | **Available stock $x_2^{s_1} + s^{s_1}$** |
|---|---|---|
| *Lowest* | 391.00 | 390.91 |
| *Very low* | 386.25 | 386.17 |
| *Low* | 440.75 | 441.07 |
| *Below average* | 504.50 | 504.93 |
| *Average* | 568.00 | 568.36 |
| *Above average* | 632.00 | 631.69 |
| *High* | 695.00 | 694.79 |
| *Very high* | 744.5 | 757.78 |
| *Highest* | 821.0 | 820.69 |

Table 4.3: Comparison of available stock and the likeliest future demand

the process, they can sell off the whole stock in the long run, even if both the demand realizations are the lowest possible. This is a strategy that hedges well even against the worst-case scenario.

Let us inspect the vendors actions after observing a demand scenario $s_1$. First important thing to notice is, that the vendor satisfies the whole first stage demand under each scenario, keeping only what they can not sell at the moment. The updated selling prices are generally decreasing with increasing first stage demand (except for the lowest scenario, where the vendor might not even be able to sell the remaining stock $s^{s_1}$ if they set the price too high). We could interpret this in a way, that when the vendor observes a high demand, they have less remaining product on stock before the second stage, so they set a lower price to attract more customers and order more product. We can also see, that when lower demand scenarios happen, the price is raised from the initial one, and when the demand is high, the price is lowered. To understand how the vendor chooses the second stage ordering amounts $x_2^{s_1}$, let us look at Table 4.3. For each first stage demand scenario $s_1$ we looked at the updated selling price $p_2^{s_1}$ and projected the possible demand scenarios along with their conditional probabilities. Then we looked at the most likely second stage demand scenario, which is always the same as in the first stage (but with respect to a new price level), which can be deduced from the matrix $\Pi$, where the largest probabilities are always on the diagonal. Then we compared it to the available stock, planned by the vendor before the second selling period, namely $x_2^{s_1} + s^{s_1}$. These values are compared in the second and third columns of the Table 4.3 and as we can see, they are basically the same. What this means is, that after the first selling period passes and the vendor chooses a new price, they find out the most likely future demand and refill the stock to that exact amount. The unsold product after the second period becomes worthless, so the vendor wants the stock amount which exactly satisfies the likeliest future demand. As for the third stage variables $h_2^{s_1,s_2}$, there is no need to show all their values, since they are always equal to the minimum of available stock and second stage demand. In other words, there is no reason for the vendor to keep anything on stock after the second selling period, so they sell what they can.

To summarize the strategy, the vendor first picks a price such that the lowest

demand is half the highest demand, to be able to sell everything even in the worst case. Then they order the highest scenario, so they can satisfy the first demand at all times. After that, they set a new price based on the amount of stock left, find out the likeliest future demand and refill the stock to this exact value. No matter what happens then, the vendor sells everything they can in the last stage and the unsold amount is discarded.

Now we will incorporate the advertisement choice. Same as in the previous chapter, we will consider the effect $\rho = 0.2$, i.e. 20% demand raise. Previously we computed the maximal advertisement price $c_A$, such that it is still profitable to buy it. We will do this again, but for various values the decay parameter $\beta$. Lower values of $\beta$ indicate a quicker decay of the advertisement, so the turnover prices should be lower (by turnover price we mean the advertisement cost where it stops being profitable). The results can be seen in Figure 4.3, which is the direct extension of Figure 3.4 from the previous chapter. We include the turnover prices for different decay parameters, which are summarized in Table 4.4. As we can see, the profitability of advertisement heavily depends on how much its effect withers in time. In the current setting, if the advertisement cost is lower than 696\$, it is profitable no matter the decay speed, and if higher than 1389\$, it does not make sense to buy it at all (only if we omitted the assumption that $\beta \leq 1$, which ensures that the effect gets weaker in time).

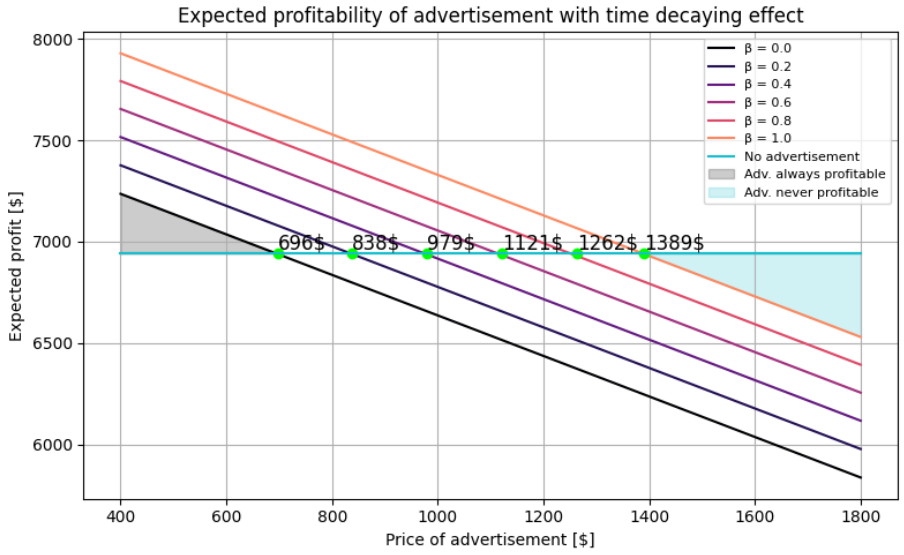


Figure 4.3: Profitability of advertisement in the three-stage model

| $\beta$ | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1 |
|---|---|---|---|---|---|---|
| **Turnover price** | 696\$ | 838\$ | 979\$ | 1121\$ | 1262\$ | 1389\$ |

Table 4.4: Turnover prices of advertisement for different rates of effect decay

# 5. Portfolio optimization with Conditional Value at Risk

Endogenous randomness in stochastic programming deals with cases, where the decision maker can influence or determine the shape of the underlying probability distribution. As far as portfolio optimization problems go, the randomness usually lies within the returns or losses of the assets of interest. Since we do not expect a normal investor to influence the asset performance in any notable way, it is usually considered a source of exogenous randomness. What if the investor had a fairly large capital and wanted to invest into smaller companies or startups? Investing a lot of money into such companies could give the investor a substantial fragment of its stocks. For example, such large investments can influence the market movement by raising the demand and showing confidence in the company, and by owning a large fragment of the stocks, the investor can even make changes in how the company operates and conduct large investments into innovation and restructurization, thus influencing the profitability and risk profile of the stock. The question is, how to model this mathematically. We will work under the mean-risk portfolio optimization framework, assuming that the investor is primarily interested in the trade-off between expected profit and risk, instead of personal feelings for the company for example. The asset returns will be modelled by the simple rate of return $r_t = \frac{P_t - P_{t-1}}{P_{t-1}}$, i.e. the percentage change of price after some time period, by losses we will mean $-r_t$. The risk modelling part is more nuanced, since there is no universal notion for it. This is where risk measures come in handy, functionals which assign a real number to a random variable, generally interpreted as a quantificiation of risk associated with the random variable. In this context, the random variable is the return or loss of some asset and the risk measure serves for comparison of different investment options. In this chapter we will utilize the arguably most famous measure **Conditional Value at Risk** (also known as Expected shortfall), which we denote by $CVaR_\alpha$ for a chosen confidence level $\alpha \in (0, 1)$ (the value $\alpha = 1$ can also be used after slight redefinition of the measure, but we will not include it here). Let $\boldsymbol{\xi} : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ be a random variable with finite moments and a quantile function $q : [0, 1] \to \mathbb{R}$. During this chapter, the random variable $\boldsymbol{\xi}$ will represent the random **loss** of an investment during some fixed time period, $-\boldsymbol{\xi}$ will be the random return.

**Definition 6.** *(Pflug (2000)) For a given confidence level $\alpha \in (0, 1)$, the risk measure **Conditional Value at Risk** of the loss variable $\boldsymbol{\xi} \in \mathcal{L}_\infty(\Omega, \mathcal{A}, \mathcal{P}_\Omega)$ is defined as*

$$CVaR_\alpha(\boldsymbol{\xi}) = \frac{1}{1-\alpha} \cdot \int_\alpha^1 q(p)\, dp.$$

The level $\alpha$ is usually taken to be 0.95 or $0,99$. In the case of 0.95, we can interpret $CVaR_{0.95}$ as an average of the 5% worst possible losses. It contains information about the tail risk of the loss distribution as it is essentially the expected value of the upper $1 - \alpha$ tail. $CVaR$ is one of the most fundamental risk measures

and it satisfies the coherence property, which is a set of conditions defining what a reasonable risk measure should have. In order to compute the measure, we need to use the following important formula (see Pflug (2000)).

$$CVaR_\alpha(\boldsymbol{\xi}) = \min_{a\in\mathbb{R}} \left( a + \frac{1}{1-\alpha} \cdot \mathbb{E}[\boldsymbol{\xi} - a]^+ \right).$$

*Remark.* Note that the set of optimal solutions to this problem is $[VaR_\alpha(\boldsymbol{\xi}), VaR^\alpha(\boldsymbol{\xi})]$, i.e. the values between the lower and upper Value at Risk. For continuous distributions of $\boldsymbol{\xi}$ it holds that $a^* = VaR_\alpha(\boldsymbol{\xi}) = q(1-\alpha)$, i.e. the $(1-\alpha)$ quantile. Value at Risk is another fundamental risk measure, which is unfortunately not coherent and does not include enough information about the full upper tail of the loss distribution.

## 5.1 Basic model formulation

Let us have $d \in \mathbb{N}$ assets and we want to construct a portfolio of them. Note that we will generally work with daily losses. A portfolio will be understood as a vector of weights

$$\lambda \in \Lambda = \left\{ \lambda \in \mathbb{R}^d_+ : \sum_{i=1}^{d} \lambda_i = 1 \right\}, \tag{5.1}$$

where $\lambda_i$ denotes the fraction of wealth invested into the $i$-th stock. These fractions will serve as decision variables in our model and since we assume the existence of endogenous randomness, they will have an effect on the loss distribution. Note that we do not allow for short positions by constraining the weights to non-negative real numbers. For each feasible portfolio $\lambda \in \Lambda$, the asset losses are stored in the random vector $\boldsymbol{\xi}(\lambda) = (\boldsymbol{\xi_1}(\lambda), ..., \boldsymbol{\xi_d}(\lambda))^T : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}^d)$ with a distribution $\mathcal{P}(\lambda)$, where $\boldsymbol{\xi_i}(\lambda)$ is the random daily loss of the $i$-th asset under portfolio choice $\lambda$. We assume all present variables to have a finite expectation. The total portfolio loss can be computed as $\lambda^T\boldsymbol{\xi}(\lambda) = \sum_{i=1}^d \lambda_i \cdot \boldsymbol{\xi_i}(\lambda)$. The general CVaR model with endogenous randomness is then formulated as

$$\begin{aligned} \min_{\lambda\in\Lambda} \; & CVaR_\alpha\big(\lambda^T\boldsymbol{\xi}(\lambda)\big) \\ \text{s.t.} \; & -\mathbb{E}[\lambda^T\boldsymbol{\xi}(\lambda)] \geq r_{min}, \end{aligned}$$

or equivalently

$$\begin{aligned} \min_{\lambda\in\Lambda, a\in\mathbb{R}} \; & \big(a + \frac{1}{1-\alpha} \cdot \mathbb{E}[\lambda^T\boldsymbol{\xi}(\lambda) - a]^+\big) \\ \text{s.t.} \; & -\mathbb{E}[\lambda^T\boldsymbol{\xi}(\lambda)] \geq r_{min}. \end{aligned}$$

where $r_{min}$ is the minimal expected return we want the portfolio to have. Since the expected value is linear, we can also write $\mathbb{E}[\lambda^T\boldsymbol{\xi}(\lambda)] = \sum_{i=1}^d \lambda_i \cdot \mathbb{E}[\boldsymbol{\xi_i}(\lambda)]$.

In other words, we prescribe a minimal expected return and search for the least risky portfolio which achieves it. A similar formulation of a decision dependent stochastic programming model can be seen in Kopa (2018), but using expected return maximization and stochastic dominance constraints.

The two most important questions are, how to obtain the loss distributions of our stocks and how to model their dependence on the portfolio structure. We will be using the empirical distribution of a basic decision independent random vector $\boldsymbol{\xi} : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}^d)$ estimated from historical data. This means that we will obtain $S \in \mathbb{N}$ observed loss scenarios and assign them equal probabilities $\frac{1}{S}$. This distribution is decision independent and we can store its scenarios inside a matrix $\mathbb{L} \in \mathbb{R}^{S \times d}$, where the $i$-th row $\xi^i \in \mathbb{R}^d$ is the $i$-th observation of the basic loss vector. We will assume a case, where the decision dependent distributions $\boldsymbol{\xi}(\lambda)$ only have transformed scenarios from the matrix $\mathbb{L}$, but their probabilities remain fixed. This is the type of dependence mentioned in Section 2.5.2. The scenario matrix for a specific portfolio $\lambda \in \Lambda$ will be denoted $\mathbb{L}(\lambda) \in \mathbb{R}^{S \times d}$ and the $i$-th loss scenario $\xi^i(\lambda)$. The positive part expectation in the objective function can then be rewritten as

$$\mathbb{E}[\lambda^T \boldsymbol{\xi}(\lambda) - a]^+ = \sum_{i=1}^{S} \frac{1}{S} \cdot \left[\lambda^T \xi^i(\lambda) - a\right]^+,$$

and using the classical optimization trick for the positive part function, we can reformulate our problem into a linear program. For the expected return constraint we decided to annualize the daily expected returns, so the parameter $r_{min}$ can be on a more intuitive scale (daily returns are very small numbers), but the rest is still in the scope of daily losses. The annualization in the return constraint does not change the model structure, only the scale of the parameter $r_{min}$. When we consider the standard 252 trading days in a year, the expected annualized return for a portfolio $\lambda \in \Lambda$ will be denoted $r(\lambda) = -252 \cdot \frac{1}{S} \sum_{i=1}^{S} \xi^i(\lambda)$. Now we can formulate the general model of interest (5.2).

$$\begin{aligned} \min_{\lambda \in \Lambda, a \in \mathbb{R}, z_1, \dots, z_m \in \mathbb{R}} & \Big(a + \frac{1}{1-\alpha} \cdot \sum_{i=1}^{S} \frac{1}{S} \cdot z_i\Big) \\ \text{s.t. } & \lambda^T r(\lambda) \geq r_{min} \\ & z_i \geq \lambda^T \xi^i(\lambda) - a, \ i = 1, \dots, S \\ & z_i \geq 0, \ i = 1, \dots, S. \end{aligned} \tag{5.2}$$

## 5.2 Problem setting

The specific shape of the dependence requires information about the stocks at hand. Remember we have $d \in \mathbb{N}$ stocks. We will denote by $M_1, \dots, M_d > 0$ the market capitalizations of the corresponding companies at the time of investment.

Market capitalization will be understood as the total value of stocks for a company, i.e. the number of stocks times the price of the stock. As we mentioned earlier, we are demonstrating endogenous randomness using the situation of a large investor and small companies. This means that the investors budget $B > 0$ can be large enough to be able to buy a considerable fraction of stocks for at least one of the considered assets. We will utilize the important modelling technique from Section 2.3, where we divide the set of feasible solutions into several disjoint parts and define a new probability distribution on each one. In this model, two arbitrary thresholds will be considered for each stock. The first threshold is hit, when the investor obtains at least 10% of the company market capitalization, then we will call them a **significant owner**. The second threshold is hit, when the investor obtains at least 50%, then we will call them a **majority owner**. We will assume that a significant owner can boost the market confidence in the stock and create an incentive for other investors to buy, whereas the majority owner can make significant structural changes inside the company and help it invest into new technologies and other important factors. The exact form of the effects will be specified later. To summarize, we assume that the investor has a positive influence on the stocks performance, and the more they own, the better. Note that this is a demonstrative example and is not meant to be perfectly realistic.

It was stated in the previous section that we have a basic loss matrix $\mathbb{L}^{S\times d}$ of $S$ equiprobable loss scenarios, observed in some past time horizon. The basic distribution is then the empirical distribution of these past observations. The endogenous randomness will be incorporated by suitable transformations of the basic scenario matrix for different ownership levels. To define the model, let us first obtain the desired quantities.

Let us consider an arbitrary stock $i \in \{1, ..., d\}$ and its portfolio weight $\lambda_i \in [0, 1]$. We know that the present market capitalization is $M_i$ and the budget is $B$. We need to compute the ownership thresholds in terms of portfolio weights. The amount of money invested into the stock is equal to $\lambda_i \cdot B$. The basic threshold will always be considered $T_i^0 = 0$. The weight $\lambda_i$ needed to hit the 10% threshold solves the equation

$$\lambda_i \cdot B = 0.1 \cdot M_i.$$

The threshold will then be denoted $T_i^1 = 0.1 \cdot \frac{M_i}{B} > T_i^0$.

Analogically for the 50% threshold

$$\lambda_i \cdot B = 0.5 \cdot M_i.$$

The threshold will be denoted $T_i^2 = 0.5 \cdot \frac{M_i}{B} > T_i^1$.

The last threshold provides the maximum weight which can be invested into the stock. The investor obviously can not buy more than the whole market capitalization. The last threshold then solves the equation

$$\lambda_i \cdot B = M_i.$$

The last threshold will be denoted $T_i^3 = \frac{M_i}{B} > T_i^2$ for each stock $i$.

From now on, we will assume for simplicity that $T_i^3 > 1 \ \forall i = 1, ..., d$, i.e. none of the market capitalizations can be completely covered by the investors budget. The consequence of this assumption is, that we can still set the upper bound for each weight to one. To summarize, there are three options for each asset, which can be indexed by the set $\{0, 1, 2\}$. 0 stands for the case when the investor owns less than 10% of the stocks, 1 stands for the case where they own at least 10% but less than 50%, and lastly ,2 stands for the case where they own at least 50% (but of course not more than 100%). We have $d$ stocks and three options for each one, which means a total of $3^d$ ownership combinations, which can be indexed by elements from the set $K = \{0, 1, 2\}^n$. Remember that the set of all possible weight vectors is $\Lambda$ (i.e. non-negative and summing to one), as mentioned in (5.1). The set of feasible weight vectors which also satisfy a combination of thresholds $k = (k_1, ..., k_n) \in K$ will be denoted by $\Lambda(k) = \Lambda(k_1, ..., k_n)$. The important thing to notice is, that under the previous assumption, these sets create a disjoint partition of the set $\Lambda$, and we can create different loss distributions for each combination of thresholds and separate the whole problem into $3^d$ smaller subproblems, while covering the whole $\Lambda$ (it is important to assume that the upper thresholds are larger than one and the weights can attain a value one, otherwise the sets $\Lambda(k)$ would not cover the whole $\Lambda$). See that this is the exact situation as in Section 2.3. The set $\Lambda(k)$ for each $k \in K$ is defined by the ownership constraints in the following way.

$$\Lambda(k) = \Lambda(k_1, ..., k_n) = \Big\{\lambda \in \Lambda : \lambda_i \in [T_i^{k_i}, T_i^{k_i+1}) \ \forall i \in \{1, ..., d\}\Big\}. \tag{5.3}$$

Strictly speaking, the last segment $[T_i^2, T_i^3)$ for each stock should be closed from the right, but since the upper bound is always larger than one, we do not need to include it, because it does not belong to the set $\Lambda$ anyway. It is vital to realize, that some of the thresholds can lie outside the interval $[0, 1]$ (not just the last ones), this happens if the investor does not have enough funds to hit them. However, this does not violate the definition of the feasibility sets, since they also contain the bounds on non-negativity and summing up to one, as they are subsets of $\Lambda$. This just means that some of the sets $\Lambda(k)$ will be empty, possibly most of them, since the investor will not be able to fulfill all the constraints (mostly because of insufficient budget). We chose this model in this way, because it reacts to the changes of budget $B$ and when the budget is raised, more ownership options become feasible and affordable, because the thresholds are lower. To understand the definition, let us take a quick example of $n = 2$ assets and $k = (0, 2)$. Then the set $\Lambda(0, 2)$ contains weight pairs, where the investor buys less than 10% of the first company stocks and more than 50% of the second.

For the final part of the model formulation, we need to set the scenario matrices on each of the sets $\Lambda(k)$. Generally, for a vector of weights $\lambda$, we denoted the decision dependent scenario matrix as $\mathbb{L}(\lambda)$. Now we only consider a finite number of distributions, induced by the partitioning $\Lambda(k)$, $k = K$ of the feasible weights set $\Lambda$. For each of these sets we want a separate scenario matrix $\mathbb{L}(k) = \mathbb{L}(k_1, ..., k_n)$. In other words, it will hold $\mathbb{L}(\lambda) = \mathbb{L}(k)$ for all $\lambda \in \Lambda(k)$. We assumed earlier, that the investor has a positive influence on the stock performance. For each

stock we will create a translation factor, depending on its performance, and move the loss scenarios down by some multiple of it, depending on the ownership level. Denote the elements of the basic scenario matrix by $\mathbb{L}_{j,i} = \xi_i^j$, $j = 1, ..., S$ and $i = 1, ..., d$, i.e. the $j$-th loss scenario of stock $i$. For an arbitrary stock $i$, the translation factor will be computed as

$$\tau_i = \left| \frac{1}{S} \cdot \sum_{j=1}^{S} \mathbb{L}_{j,i} \right| = \left| \frac{1}{S} \cdot \sum_{j=1}^{S} \xi_i^j \right|,$$

i.e. the absolute value of the mean daily loss for the stock. The absolute value is there because the factor is supposed to be positive at all times, in order to lower the loss scenarios. Then we define three multiplicative constants for each $k = 0, 1, 2$ denoted $\kappa_0, \kappa_1, \kappa_2 \geq 0$ and we set $\kappa_0 = 0$. For each combination $k = (k_1, ..., k_n) \in K$ is the new scenario matrix defined as

$$\mathbb{L}(k) = \begin{bmatrix} \mathbb{L}_{1,1} - \kappa_{k_1} \cdot \tau_1 & \cdots & \mathbb{L}_{1,n} - \kappa_{k_n} \cdot \tau_n \\ \mathbb{L}_{2,1} - \kappa_{k_1} \cdot \tau_1 & \cdots & \mathbb{L}_{2,n} - \kappa_{k_n} \cdot \tau_n \\ \vdots & \ddots & \vdots \\ \mathbb{L}_{m-1,1} - \kappa_{k_1} \cdot \tau_1 & \cdots & \mathbb{L}_{m-1,n} - \kappa_{k_n} \cdot \tau_n \\ \mathbb{L}_{m,1} - \kappa_{k_1} \cdot \tau_1 & \cdots & \mathbb{L}_{m,n} - \kappa_{k_n} \cdot \tau_n \end{bmatrix}.$$

We basically increase the expected performance of $i$-th stock by $100 \cdot \kappa_{k_1}\%$ for an ownership level $k_i$. Previously we set $\kappa_0 = 0$ so that the basic scenarios for a specific stock remain unchanged if the investor owns less than 10% of the market capitalization. We did not want to move all the columns by the same number, so we created the translation factor to scale the change of scenarios for each stocks average loss. For normal sized budgets of small investors, this will result in the classical $CVaR$ optimization problem, since all the non-trivial thresholds will be far above one. The changes in loss scenarios occur only when the budget is large enough so that some of the ownership thresholds fall below one and are therefore affordable. Denote $K^* = \left\{ k \in K : \Lambda(k) \neq \emptyset \right\}$, i.e. all the ownership combinations which have a feasible solution (meaning that the investor can distribute the whole budget in a way that satisfies the threshold constraints). Note that there is a total of $3^d$ subproblems, which is a number that can explode for a large amount of stocks. However, in principle, many of them will have no feasible solutions, since the sets $\Lambda(k)$ are often empty (depending on the budget size). In order to save resources, we can create a condition to cut combinations which are sure to result in an empty feasibility set. Two situations can happen, which ensure there is no solution for a specific constraint combination $k \in K$. First situation occurs, when the investor does not have enough funds to satisfy the constraints (5.3). Second situation occurs, when even after buying the maximum stock amounts allowed by the constraints, the investor still has unallocated budget left. To summarize, we know that if $k \in K^*$ both the conditions (5.4) must hold.

$$\begin{aligned} &\sum_{i=1}^{d} T_i^{k_i} \leq 1, \\ &\sum_{i=1}^{d} T_i^{k_i+1} > 1. \end{aligned} \tag{5.4}$$

The first condition ensures, that the investor can afford to buy at least the minimum amount prescribed by the constraints. The second condition ensures, that it is possible to distribute the whole budget while satisfying the constraints. If it does not hold, the investor can buy the maximum possible amount and still have unspent budget left. The strict inequality is there because of the strict upper bound in the constraints (5.3). Buying the exact upper bounds $T_i^{k_i+1}$ would result in leaving the set $\Lambda(k)$. We can use condition (5.4) to filter out infeasible combinations without having to solve the subproblem.

To summarize, for each $k \in K^*$ we have the scenario matrix $\mathbb{L}(k) \in \mathbb{R}^{S\times d}$ with rows (scenarios) $\xi^i(k)$ of daily losses and a vector of annualized expected returns $r(k)$. Note that while implementing the model into a solver, we replace all the constraints $\lambda_i < T_i^{k_i+1}$ from (5.3) by $\lambda_i \leq T_i^{k_i+1} - \epsilon$ for some very small positive number $\epsilon$. The notation and logic behind the quantities is the same as in the general model (5.2) and everything is prepared to formulate the final model (5.5).

$$\begin{aligned} \min_{k\in K^*} \min_{\lambda\in\Lambda(k),a,z_1,...,z_m\in\mathbb{R}} &\left(a + \frac{1}{1-\alpha}\cdot\sum_{i=1}^{S}\frac{1}{S}\cdot z_i\right) \\ \text{s.t. } &\lambda^T r(k) \geq r_{min} \\ &z_i \geq \lambda^T\xi^i(k) - a,\ i = 1,...,S \\ &z_i \geq 0,\ i = 1,...,S, \end{aligned} \tag{5.5}$$

or after including the weight constraints explicitly,

$$\begin{aligned} \min_{k\in K^*} \min_{\lambda\in\mathbb{R}^d,a,z_1,...,z_m\in\mathbb{R}} &\left(a + \frac{1}{1-\alpha}\cdot\sum_{i=1}^{S}\frac{1}{S}\cdot z_i\right) \\ \text{s.t. } &\lambda^T r(k) \geq r_{min} \\ &\sum_{i=1}^{d}\lambda_i = 1 \\ &\lambda_i \geq 0,\ i = 1,...,d \\ &\lambda_i \in [T_i^{k_i}, T_i^{k_i+1}),\ i = 1,...,d \\ &z_i \geq \lambda^T\xi^i(k) - a,\ i = 1,...,S \\ &z_i \geq 0,\ i = 1,...,S. \end{aligned} \tag{5.6}$$

We can now proceed to a numerical example for real stock market data.

## 5.3 Numerical example

Let us now set up the important model parameters. We will work under the commonly used confidence level $\alpha = 0.95$.. As for the dependency parameters, we set $\kappa_0 = 0$, $\kappa_1 = 0.1$, $\kappa_2 = 0.25$, which means that the loss scenarios remain unchanged if the investor buys less than a tenth of the company stocks, a significant owner can increase the performance by 10% and a majority owner by 25%. These values are best estimated from data about market reactions to large investments, but obtaining such data is not simple. A possible way of obtaining these values could be through expert opinion.

Since we need to estimate the Conditional Value at Risk by a sample estimate, we would like at least 30 observations at the $1-\alpha = 0.05$ tail (to apply the law of large numbers), which means around 600 observation in total. Startup companies would be the best candidates, but there is not enough observations to estimate from, whereas older companies with a small capitalization are often bankrupted. We chose $d = 5$ viable stocks using the Yahoo Finance stock screener, filtering for companies with a small market capitalization. We imported the daily adjusted closing prices from January 1st 2022 to June 1st 2024, which (after computing the daily losses) made for a total of $S = 605$ observations. The stocks will be referred to by their tickers and relevant information about them are summarized in Table 5.1. Note that the market capitalizations in the Table were observed on July 17th 2024, which we consider to be the day of investment in this example. They obviously vary with the stock price, so the updated values should be used when applying the model at a different time. The Table 5.2 contains the estimated yearly returns for each stock under each of the three ownership options. The first is estimated from the data, the others are transformed using the method (5.2) from the previous section. Note that those are the values $r(k)$ later entering the model, but in percentage terms.

| **Stock ticker** | **Sector of interest** | **Market cap. [mil. USD]** |
|---|---|---|
| **ALOT** | Computer Hardware / Technology | 115.78 |
| **CAAS** | Auto Parts / Consumer Cyclical | 114.67 |
| **HMENF** | Oil & Gas / Energy | 122.13 |
| **INTT** | Semiconductor Equipment / Technology | 119.53 |
| **ISSC** | Aerospace & Defense / Industrials | 111.25 |

Table 5.1: Overview of the stocks of interest

| **Stock ticker** | **Basic return** | **Significant ownership** | **Majority ownership** |
|---|---|---|---|
| **ALOT** | 18.55% | 20.41% | 23.19% |
| **CAAS** | 34.99% | 38.49% | 43.74% |
| **HMENF** | 34.70% | 38.17% | 43.37% |
| **INTT** | 4.73% | 5.21% | 5.92% |
| **ISSC** | 9.05% | 9.95% | 11.31% |

Table 5.2: Estimated and adjusted annualized expected returns

A portfolio is said to be **efficient**, when there is no other portfolio with a higher or equal expected return and lower or equal risk (and at least one of the inequalities strict). Mean-risk portfolio optimization is essentially a multi-objective programming problem, which we deal with by adding one of the objectives (expected return) in the constraints, using a minimal return parameter. For each attainable level of $r_{min}$ is the optimal portfolio efficient (if the solution is unique). Plotting these efficient portfolios on a two dimensional graph creates the efficient frontier, which summarizes the risk-return trade-off for different levels of minimal return. The minimal optimal expected return (the starting point of the frontier) will be obtained by solving the optimization problem (5.6) without the minimal return constraint, i.e. the expected yearly return of a portfolio with the absolutely minimal $CVaR$, no matter the return. The main goal of this example is to see how efficient frontiers change after introducing endogenous randomness into the model. Do they keep the same properties? In order to present the quantities on a more intuitive scale, the expected returns will be yearly in percentage terms and the Conditional Values at Risk will be daily, also in percentage terms. All plots will also include the mean-risk profile of a thousand other portfolios with weights randomly sampled from the Dirichlet distribution. For each randomly generated portfolio will the loss scenarios be adjusted to the combination of exceeded weight thresholds and the $CVaR$ and expected yearly return will be computed and plotted for comparison. Since the efficient frontier maps efficient portfolios, all randomly generated portfolios should lie on or beneath the curve, otherwise they would deny the efficiency property.

We will show efficient frontiers for investors with three different budgets, 1 million USD, 30 million USD and 90 million USD. Judging from the market capitalizations in the Table 5.1, the first budget of 1 million USD was chosen so that the investor can not afford to become a significant owner anywhere. This should result in the classical portfolio optimization model, where the investor does not have enough funds to change the behavior of the stocks at hand. In other words, it corresponds to the portfolio optimization problem with exogenous randomness. Figure 5.1 shows, that the efficient frontier has the classical non-decreasing and concave shape. Let us now explore the frontiers for higher budgets, where the investor has the means to alter the loss distribution. The medium budget of 30 million USD was chosen so that the investor can become a significant owner in an arbitrary company, but never the majority owner. The efficient frontier is shown in Figure 5.2. Finally, the large budget of 90 million USD is chosen so that the investor can become a majority owner in a company of choice, resulting in a higher number of non-empty sets $\Lambda(k)$. The efficient frontier is shown in Figure 5.3. All the three frontiers are compared in Figure 5.4.

To summarize, efficient frontier is always non-decreasing, since a higher return requires higher risk. However, we can see that with a sufficient budget, the efficient frontier is no longer the concave curve like in the classical case. On both the purple and red curves, we can see spots with rapid changes in the direction, which look like inflexion points, induced by the investor acquiring a higher ownership level for some stock. The concave shape of the classical efficient frontiers comes from the fact, that achieving a higher return gets progressively more difficult and

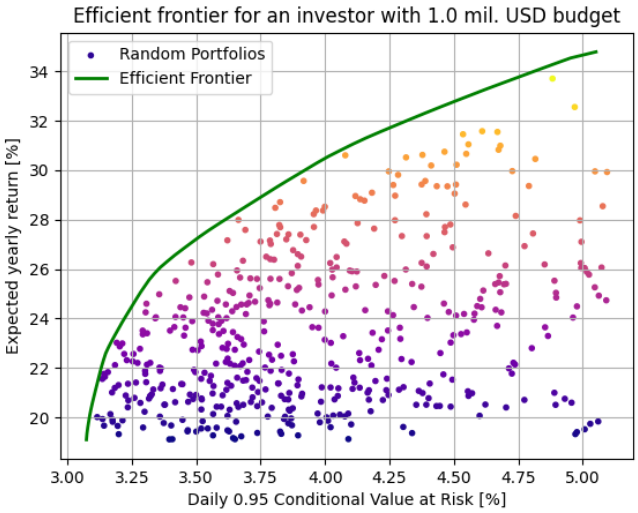


Figure 5.1: The efficient frontier for a small budget

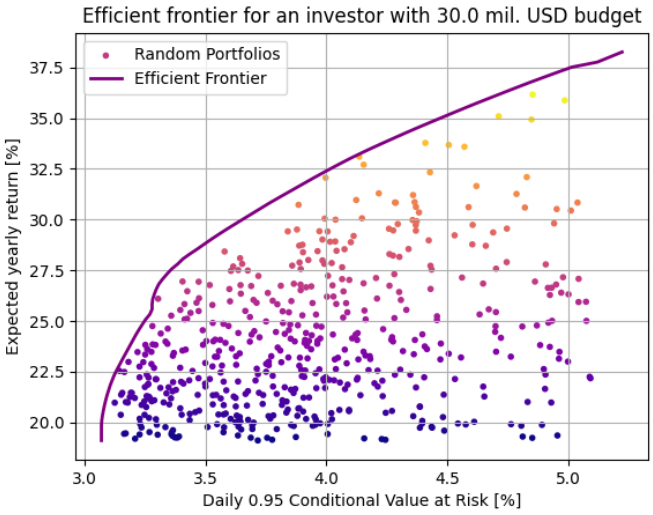


Figure 5.2: The efficient frontier for a medium budget

requires more and more investments into volatile stocks. As we can see, the concavity property can be broken in situations, where the expected return increases rapidly, without the need to raise the risk that much.

The expected return of the portfolio with the absolute smallest CVaR is 19.11%. To show at least some specific optimal portfolio weights, let us pick two arbitrary values for the minimal expected return $r_{min}$, namely 25% and 30%, and compute portfolios for our three budget levels. The results are shown Table 5.3. Note that we are dealing with small numbers here and we could lose too much information if we rounded to the usual two decimal places, as in the rest of the thesis, so we will use three digits here.

We can consult the Table 5.2 showing the expected annualized returns of the stocks. For the higher required return of 30%, we can see that a lot more has to

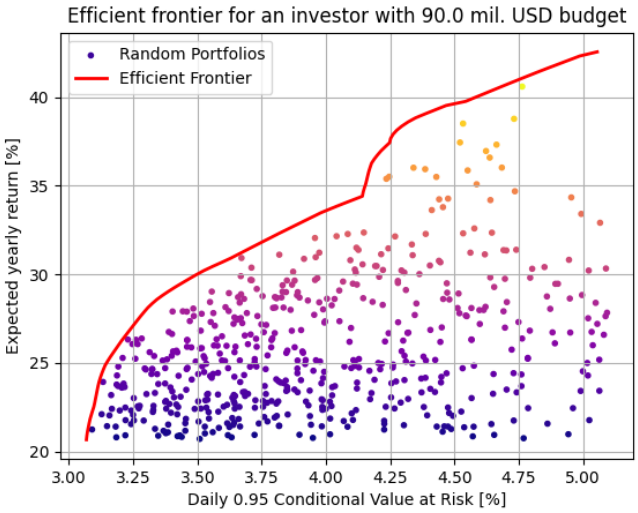


Figure 5.3: The efficient frontier for a large budget

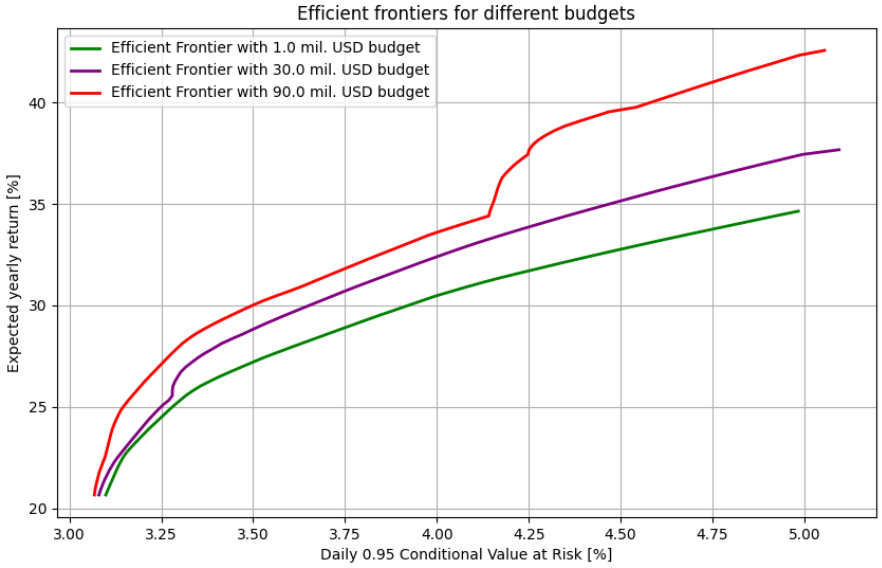


Figure 5.4: The comparison of efficient frontiers

be invested into CAAS and HMENF, which have the highest expected returns. The least profitable efficient portfolio has an expected return of 19.11%, which is a lot more than the expected returns of INTT and ISSC. Those are the least profitable stocks but their risk is obviously not that good either, which is why INTT is rarely even present in the portfolios. We can also see that for larger budgets the investor puts less into the most profitable stocks CAAS and HMENF, since they are risky and it is easier to satisfy the minimal return without them, because of the increased ownership levels. On the contrary, the ISSC stock has an increasing optimal weight with increasing budget. It is not that profitable but has a low risk, therefore with a higher budget the investor can afford to put more in it, while still satisfying the minimal return.

For the minimal expected return of 25%, most funds are always put into the stocks HMENF (with the second highest return) and ALOT (third highest return). CAAS and ISSC are similarly represented, the first having a larger return and the other having a lower risk. When the budget is large, the investor can afford to put more into the less profitable ISSC. The stock INTT is almost not represented. For the minimal return of 30%, most funds are again put into the stocks HMENF and ALOT. CAAS and ISSC are again similarly represented, but in this case the investor needs to put more into CAAS, since the minimal return requirement is larger. The stock INTT is not represented at all. To summarize, we can see that having a larger budget allows the investor to raise the expected returns of some of the stocks, which is why they can invest more into the less risky assets, while still satisfying the minimal return requirement. The portfolios for the lowest budget are essentially the portfolios we would obtain from the classical exogenous problem. By comparing them to the ones corresponding to the highest budget, we can see that the incorporation of endogenous randomness has a significant effect on portfolio choice and the differences in weights are notable.

| **Budget** | $r_{min}$ | **ALOT** | **CAAS** | **HMENF** | **INTT** | **ISSC** |
|---|---|---|---|---|---|---|
| 1 mil. | 0.25 | 0.320 | 0.116 | 0.387 | 0.004 | 0.173 |
| 30 mil. | 0.25 | 0.386 | 0.102 | 0.348 | 0.000 | 0.164 |
| 90 mil. | 0.25 | 0.281 | 0.161 | 0.276 | 0.050 | 0.232 |
| 1 mil. | 0.3 | 0.258 | 0.199 | 0.520 | 0.000 | 0.023 |
| 30 mil. | 0.3 | 0.237 | 0.169 | 0.491 | 0.000 | 0.102 |
| 90 mil. | 0.3 | 0.267 | 0.168 | 0.442 | 0.000 | 0.124 |

Table 5.3: Budget dependent optimal portfolios for two values of the minimal expected return

# Conclusion

In this thesis we studied a very interesting subtopic of stochastic optimization. The assumption that the underlying probability distribution is not dependent on the decisions was challenged and removed. The goal of this thesis was to explain the concept of endogenous randomness in stochastic programming and think about model types which are computationally and mathematically tractable.

In the first chapter we provided an overview of relevant stochastic optimization topics, which might later be generalized and expanded into the new framework. We provided a general uncertain form of the model and discussed various methods of deterministic reformulation. A significant topic was also the multi-stage programming, including the basics of scenario trees and nonanticipativity constraints.

The second chapter was the core of this thesis, since it was dedicated to laying the groundwork for the modelling of endogenous randomness in stochastic optimization. We tried to create a coherent overview of the basic methods and formulations, using a unified notation and mathematical background, which is not frequently found in literature on this topic. The most notable cases were: division of the feasibility set into subsets generating a single common distribution, decisions affecting the parameters of a family of distributions, decisions affecting the scenario or probabilistic structure for finitely many scenarios, and decisions affecting the dependence structure or marginal distributions. Later we removed the assumption of unique linkage between decisions and distributions and presented methods used for dealing with a set-valued dependence, mainly the robust approach and aggregation using mixture distributions. Lastly, we talked a little about the stability of optimal solutions and gave a brief introduction to multi-stage optimization with endogenous randomness.

In the third chapter we presented the well known newsvendor model and provided a custom extension for continuous price selection and advertisement choice. We have set a parametric dependence on a continuous uniform distribution and derived the deterministic reformulation. We presented the model on a numerical example using ad hoc parameters and functions, provided plots and studied the profitability of advertisement under varying cost.

In the fourth chapter we changed the distribution of demand and formulated a three-stage extension for the newsvendor model, including advertisement choice and recourse actions based on the observed demand. We pursued the tractable case of a fixed shape of the scenario tree and demonstrated the model on a numerical example using a predefined probabilistic structure. Then we provided plots and studied the profitability of advertisement under different rates of effect decay.

In the fifth chapter we expanded the classical CVaR portfolio optimization problem to a case where large investments can alter the loss distribution of the assets.

The model reacted to budget by setting up ownership thresholds induced by the obtained portions of the companies market capitalization. The model was demonstrated on a numerical example for 5 small cap stocks and optimal portfolios were computed. Most importantly, efficient frontiers were plotted for three budget levels, where the first budget was too small to alter the loss distributions, and the two larger budgets progressively changed the classical concave shape of the curve and added inflexion points.

The main contribution of this thesis lies within the provided theoretical framework, which contains various possible types of dependence. Another contribution are the three models from chapters 3-5, on which we tried to demonstrate different methods from chapter 2. The first model was an example of the distribution having a fixed parametric family. The second example was the multi-stage case with changing scenarios and conditional probabilities. Lastly, the third example demonstrated the case of finite partition of the feasibility set using various distributions, and also the altering of a set of scenarios with fixed probabilities.

# Bibliography


Basciftci, B., Ahmed, S., and Shen, S. (2021). Distributionally robust facility location problem under decision-dependent stochastic demand. *European Journal of Operational Research*, 292(2):pages 548–561.

Dupačová, J. (2006). Optimization under exogenous and endogenous uncertainty. In *Mathematical Methods in Economics*, pages 131–136.

Goel, V. and Grossmann, I. E. (2004). A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves. *Computers and chemical engineering*, 28(8):pages 1409–1429.

Hrabec, D., Haugen, K., and Popela, P. (2017). The newsvendor problem with advertising: an overview with extensions. *Review of Managerial Science*, 11:pages 767–787.

Hrabec, D., Popela, P., Novotný, J., Haugen, K., and Olstad, A. (2012). A note on the newsvendor problem with pricing. In *18th International Conference on Soft Computing, MENDEL*, pages 410–415.

Jonsbråten, T. W., Wets, R. J.-B., and Woodruff, D. L. (1998). A class of stochastic programs with decision dependent random elements. *Annals of Operations Research*, 82(0):pages 83–106.

Kopa, M. (2018). Endogenous randomness and first–order stochastic dominance in portfolio optimization. In *Mathematical Methods in Economics*, pages 246–251.

Kopa, M. (2024). Stochastic programs with decision dependent randomness - contamination and stress testing. Working paper.

Kopa, M. and Rusý, T. (2021). A decision-dependent randomness stochastic program for asset-liability management model with a pricing decision. *Annals of Operations Research*, 299(1):pages 241–271.

Kopa, M. and Rusý, T. (2023). Robustness of stochastic programs with endogenous randomness via contamination. *European Journal of Operational Research*, 305(3):pages 1259–1272.

Luo, F. and Mehrotra, S. (2020). Distributionally robust optimization with decision dependent ambiguity sets. *Optimization Letters*, 14(8):pages 2565–2594.

Pflug, G. (1990). On-line optimization of simulated markovian processes. *Mathematics of Operations Research*, 15(3):pages 381–395.

Pflug, G. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. *Probabilistic constrained optimization: Methodology and applications*, pages 272–281.

Ruppert, D. and Matteson, D. S. (2011). *Statistics and data analysis for financial engineering*, volume 13. Springer.

Shapiro, A., Dentcheva, D., and Ruszczynski, A. (2009). *Lectures on stochastic programming: modeling and theory*. Society for Industrial and Applied Mathematics.

# List of Figures

# List of Tables