Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Jitka Čížková

Očekávaná hodnota informace ve stochastickém programování

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Prof. RNDr. Jitka Dupačová, DrSc.
Studijní program: Matematika
Studijní obor: Pravděpodobnost, matematická statistika a ekonometrie
Studijní plán: Ekonometrie

Děkuji paní Prof. RNDr. Jitce Dupačové, DrSc. za mnoho cenných rad a vytrvalou pomoc při psaní diplomové práce.

Prohlašuji, že jsem svou diplomovou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 4. dubna 2006 Jitka Čížková

Charles university in Prague
Faculty of Mathematics and Physics

DIPLOMA THESIS

Jitka Čížková

Expected Value of Information in Stochastic Programming

Department of Probability and Mathematical Statistics

Supervisor: Prof. RNDr. Jitka Dupačová, DrSc.
Study program: Mathematics
Study branch: Probability, Mathematical statistics and Econometry
Study plan: Econometry

Contents

Název práce: Očekávaná hodnota informace ve stochastickém programování
Autor: Jitka Čížková
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí diplomové práce: Prof. RNDr. Jitka Dupačová, DrSc.
e-mail vedoucího: dupacova@karlin.mff.cuni.cz

Abstrakt: Úlohy stochastického programování (dvoustupňové i vícestupňové) lze formulovat několika různými způsoby, které lépe či hůře využívají dostupnou informaci o budoucí realizaci náhodných parametrů. Porovnáním optimálních hodnot účelové funkce, které dostaneme při řešení rozdílně formulované úlohy při téže dostupné informaci, zjistíme, jaká je hodnota jedné z těchto formulací oproti druhé (např. *VSS*).
Úroveň zmíněné dostupné informace lze měnit částečným, resp. úplným uvolněním předpokladu neanticipativnosti, podle kterého nesmí současná rozhodnutí záviset na budoucích (neznámých) realizacích náhodných parametrů. Porovnání optimálních hodnot účelových funkcí, získaných řešením dané úlohy při nižší a vyšší úrovni dostupné informace, vede na (očekávanou) hodnotu částečné, resp. úplné informace.
V této práci uvádíme definice různých typů hodnoty informace a příbuzných hodnot souvisejících s formulací úlohy a odvození jejich vlastností (nezápornost, meze). V závěru provádíme jejich souhrnnou klasifikaci.

Klíčová slova: optimalizace, stochastické programování, hodnota informace, hodnota stochastického řešení

Title: Expected Value of Information in Stochastic Programming
Author: Jitka Čížková
Department: Department of Probability and Mathematical Statistics
Supervisor: Prof. RNDr. Jitka Dupačová, DrSc.
Supervisor's e-mail address: dupacova@karlin.mff.cuni.cz

Abstract: Stochastic problems (both two-stage and multistage) can be formulated in several different ways which utilize to various extent available information on a future realization of incorporated random parameters. When comparing optimal objective function values resulting from different formulations of the given problem with the same available information, we obtain a value of using one of these formulations rather than the other one (e.g., *VSS*).
Level of the available information can be changed by a partial or full relaxation of nonanticipativity constraints, which assure that a present decision is independent of future (unknown) realizations of random parameters. By comparing optimal objective function values gained when solving the given problem with distinct levels of available information we obtain (expected) value of partial or perfect information.
In this work we present definitions of various information value types and related values connected with the problem formulation and we derive their properties (nonnegativity, bounds). In the last part we introduce their summary classification.

Keywords: optimization, stochastic programming, value of information, value of stochastic solution

1. Introduction to mathematical programming

1.1. Introduction

In various real-life situations, we often deal with optimization problems. Their core is to find a solution which is the best one from a certain point of view (e.g. the cheapest or the most profitable one), subject to some conditions and constraints. In mathematical formulation of the problem, these constraints define a so called *feasibility set,* and the property we focus on is incorporated in an *objective function.* There can be more than one objective function (e.g. our goal can be to maximize profits and minimize ecological impact) but we will consider just one objective function problems.

The aim of a decision maker is to find a minimum or maximum (or an infimum or supremum, respectively) of the objective function subject to given constraints, i.e. his decisions must be feasible.

In practice, we often need to plan our future doing and then the objective function and/or the feasibility set are to certain extent uncertain. They can be influenced by a future development of some quantities about which a full information is not available at the time of making the decision (e.g. development of prices, interest rates or demand). In optimization, this uncertainty is modelled by use of random variables, most often with a known probability distribution.

If the decision maker could postpone his decision to a time when the development of random parameters is already known, or if he was a clairvoyant knowing the future in advance, his resulting optimal solution could only be better than that made without this information on future. Therefore, the information on the future random events is of a certain value for the decision maker. This value could be interpreted as a price which is to be paid for putting off the decision, for making a research giving this information, or as a reward for the clairvoyant who reveals the future in advance.

Given problem can also be formulated in many ways, using various approaches. The formulations can be more or less sophisticated and computationally demanding, and they usually lead to various results. For any pair of approaches and relating results the difference between the resulting optimal objective function values expresses how much it is worth doing to solve the problem using the "better" approach than the other one. Having this information, the decision maker can judge whether he shall undergo more computations in order to gain more favourable result or not.

The aim of this work is to precisely define various "types" of the value of information, to find relationships between them and to derive their estimates and properties for some specified classes of problems. In the first part, we give a brief introduction to linear and stochastic programming. Part two is devoted to various information value types in two-stage programs. Some of them are generalized in the third part, where we also derive some new information value types specific for multistage problems. All the information value types are compared and classified in the last fourth part.

1.2. Linear programming

Linear programs are the simplest problems of mathematical programming. We deal with a problem with a linear[1] objective function and with a feasibility set bounded by linear

[1] We do not distinguish between linear ($f(x) = ax$) and affine ($f(x) = ax + b$) functions.

functions. The problem reads

$$\begin{aligned}
&\min_x \sum_{j=1}^{n} c_j x_j \\
&\text{s.t. } \sum_{j=1}^{n} a_{ij} x_j \geq b_i,\ i = 1, \ldots, m, \\
&\quad\ x_j \geq 0,\ j = 1, \ldots, n,
\end{aligned} \tag{1.1}$$

where $x_j,\ j = 1, \ldots, n$ are decision variables; for $j = 1, \ldots, n$, the c_j represents (fixed and real) "unit costs" of the decision x_j; a_{ij} for $i = 1, \ldots, m$, $j = 1, \ldots, n$ and b_i for $i = 1, \ldots, m$ are given real constants which provide the constraints.
The problem (1.1) can be equivalently written in the form

$$\begin{aligned}
&\min_x\ c^T x \\
&\text{s.t. } Ax \geq b, \\
&\quad\ x \geq 0,
\end{aligned} \tag{1.2}$$

where $c = (c_1, \ldots, c_n)^T$, $x = (x_1, \ldots, x_n)^T$, $A = (a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}$, $b = (b_1, \ldots, b_m)^T$ and the inequations $Ax \geq b$ and $x \geq 0$ have the sense that $(Ax)_i = \sum_{j=1}^{n} a_{ij} x_j \geq b_i$ for $i = 1, \ldots, m$ and $x_j \geq 0$ for $j = 1, \ldots, n$, respectively. The upper indices T stand for transpositions.
Further on, any inequality of the type $u \geq v$ used for equally-dimensional real vectors u and v has the sense that $u_j \geq v_j$ for all j.
Note, that any minimization problem can be easily reformulated into an equivalent maximization problem, since for any objective function z holds $\min_x z(x) = -\max_x(-z(x))$. In this work, we will formulate all problems as the minimization ones.
Mention that we do not consider as a problem whether the constraints in a mathematical program are of the type $Ax = b$ or $Ax \geq b$, since these two types can be transformed to each other using so called slack variables. Also, we are not too interested in the problem of existence of minima we operate with; in this work we assume that all the employed minima (or maxima) exist, unless stated otherwise.

For the problem (1.2), we can construct a *dual problem*

$$\begin{aligned}
&\max_y\ b^T y \\
&\text{s.t. } A^T y \leq c, \\
&\quad\ y \geq 0.
\end{aligned} \tag{1.3}$$

In this context, the problem (1.2) is referred to as a *primal problem.*
The problems (1.2) and (1.3) have some interesting properties that make them a strongly joint pair.

The linear problems can be generalized into nonlinear ones. In the problem (1.2), the linear objective function and linear functions providing the constraints are replaced by nonlinear functions. Some types of dual problems can also be formulated but they are not of any use in this work.

1.3. Stochastic programming

1.3.1. Random variables

In this paragraph we will remind a definition of random variables and some related definitions. We will use the most common notions from probability theory and mathematical statistics without definitions.

Consider an abstract set Ω (a set of elementary events) and a σ-field $\mathscr{F}$ of its subsets. If $\Omega \equiv \mathbb{R}^k$ for a finite $k \in \mathbb{N}$, then we always equip Ω with the σ-field generated by $\mathbb{R}^k$, which is the Borel σ-field $B(\mathbb{R}^k)$. Consider a probability measure P on $(\Omega, \mathscr{F})$; the triplet $(\Omega, \mathscr{F}, P)$ is called a *probability space.*

An $\mathscr{F}$-measurable function $\xi\colon (\Omega, \mathscr{F}, P) \to \mathbb{R}^m$ is called a *real random vector* (a scalar real random variable for $m = 1$). The set of realizations of the random vector ξ is $\Xi = \{\xi(\omega) : \omega \in \Omega\} \subseteq \mathbb{R}^m$ and it is called the *support* of the random vector ξ. It is clear that $P(\{\omega : \xi(\omega) \in \Xi\}) = 1$ which we write as "$\xi \in \Xi$ almost surely."

For an arbitrary random variable ξ and for any Borel set B, there can be constructed its inverse image $\xi^{-1}(B) = \{\omega \in \Omega : \xi(\omega) \in B\}$ and a probability measure $\mu(B) = P(\xi^{-1}(B))$. This measure μ is called the *induced measure* (induced by ξ).

A probability distribution of any random variable ξ is unambiguously defined by ξ's induced measure or by its *cumulative distribution function* F_ξ which is defined as $F_\xi(x) = P(\{\omega \in \Omega : \xi(\omega) \leq x\})$, or by its probability density, if it exists. In the case of a random vector ξ we often talk about its probability distribution as the joint distribution of its components (random variables) $\xi_1, \dots, \xi_m$.

Expectation of a random variable ξ is defined as $\mathrm{E}\xi = \int_\Omega \xi dF(\xi) = \int_\Omega \xi(\omega) dP(\omega)$. For an extended real-valued random variable ξ we define $\mathrm{E}\xi = \mathrm{E}[\xi_+] - \mathrm{E}[(-\xi)_+]$, where $\xi_+(\omega) = \max\{0; \xi(\omega)\}$, and $\mathrm{E}\xi$ is not defined in the case that both $\mathrm{E}[\xi_+]$ and $\mathrm{E}[(-\xi)_+]$ are $+\infty$. In this work, unless stated otherwise, whenever we operate with expectations of random variables (e.g. $\mathrm{E}\xi$) or their transformations (e.g. $\mathrm{E}\min_{x \in K} z(x, \xi)$), we tacitly assume that the expectations exist and are finite.

Measurable functions $\xi\colon (\Omega, \mathscr{F}, P) \to \bar{\mathbb{R}}$ (or into $\bar{\mathbb{R}}^m$) are referred to as extended real valued random variables (or m-dimensional vectors). In such cases, the values $F_\xi(+\infty)$ and $F_\xi(-\infty)$ are defined as appropriate limits.

In this work, random variables (both scalar and vector) are denoted by letters of the Greek alphabet, most often by ξ and ζ. Very often we denote by the same symbol ξ a particular realization of the random variable $\xi = \xi(\omega)$ and we also often suppress its explicit dependence on the elementary events $\omega \in \Omega$. We then write $P(\xi \leq x)$ instead of $P(\{\omega \in \Omega : \xi(\omega) \leq x\})$ and the like.

Consider a space χ of all m-dimensional real random vectors defined on a certain probability space $(\Omega, \mathscr{F}, P)$. A *probability functional* is a mapping $\mathrm{F}\colon \chi \to \mathbb{R}^{\tilde{m}}$ (where, most often, $m = \tilde{m}$). For $\xi \in \chi$ the value $\mathrm{F}(\xi)$ depends on the probability distribution of the random vector ξ (but it cannot depend on any realizations of ξ since no elementary events are considered in $\mathrm{F}(\xi)$). A typical example of a probability functional is an expected value functional.

1.3.2. Stochastic problems

In stochastic programs both the objective function and the constraints can depend on some uncertain quantities which are considered as random variables. The decision maker has to find a minimum (or an infimum) of an objective function $z = z(x, \xi)$, where x is an n-dimensional vector of decision variables, ξ is a real random variable defined on a certain probability space and having a support Ξ, and $z\colon \mathbb{R}^n \times \Xi \to \mathbb{R}$, under a condition that the decision variable x belongs to a feasibility set $K(\xi)$ which depends on a particular realization of ξ. Now $z(x, \xi)$ is a random variable as well, so a probability functional F is applied there such that $\mathrm{F}[z(x, \xi)] \in \mathbb{R}$ for all x and it does not depend on a particular realization of ξ. The most general form of the stochastic problem is then

$$\min_x \mathrm{F}[z(x, \xi)]$$
$$\text{s.t. } x \in K(\xi).$$

Most often, this functional F is the expectation (denoted as E).
There are two basic assumptions that are tacitly supposed to be satisfied in this work, unless stated otherwise. Firstly, that the distribution of the random variable ξ is known (by the decision maker) in advance. Secondly, that the distribution of the random variable ξ does not depend on the decision variable x. Note that the second condition is much more realistic than the first one.

1.3.3. Time periods and stages

Real-life problems often require not only one decision, but a decision process, i.e. a sequence of decisions made in certain time points (e.g. management of a firm regularly assesses on a budget, expansion, volume of production etc.).
Consider a division of the overall time interval into several (possibly infinitely many) time periods, e.g. years. Consider a fixed t-th period. At the beginning of this period, a first stage, so called initial decision for this period, is made (e.g. manager decides how many warehouses have to be rented for the next year). Then the decision-maker alternately observes random events and reacts to them in his decisions, under conditions which are usually related to the initial decision for this period and of course subject to some other constraints (e.g. manager decides how much goods and of which type has to be transported into which warehouse to satisfy the demand, fully use the capacity, minimize transportation costs etc.). This consequent decisions are called the decisions of the second stage, third stage et cetera. Numbers of stages in particular periods can differ (e.g. according to the number of working days in the year).
In this work we divide the problems into two basic classes, to two-stage and multistage problems. In the two-stage problems the decision-maker makes his initial decision x, then he observes a random event ξ, and he reacts to it in his second stage decision $y = y(\xi) = y(x, \xi)$. In the multistage programs, the decision-maker makes an initial decision x_1, then he observes the first random event ξ_1, then he makes the second-stage decision $x_2 = x_2(\xi_1) = x_2(x_1, \xi_1)$, then he observes ξ_2, makes the third-stage decision $x_3 = x_3(x_1, \xi_1, x_2, \xi_2)$ et cetera up to the last decision $x_T = x_T(x_1, \xi_1, \ldots, x_{T-1}, \xi_{T-1})$.
For a certain class of the two-stage problems, an extensive theory regarding to their properties, searching for estimates and the related value of information is worked up. It is more difficult (and sometimes impossible) to work up a similar theory for the multistage programs. We will therefore inquire into two-stage and multistage programs in two separate chapters, and at the end we will summarize results gained in both of them.

2. Two-stage problems

2.1. Two-stage problems of linear stochastic programming

Two-stage linear programs are one of the most simple types of stochastic problems. Generally, the problem is to minimize (expected) costs or outcomes. The decision maker has to decide in two stages. At first he makes *first stage decision*. It has to be taken without full information on some (future) random events, which are modelled with help of random variables represented by a random vector ξ. The decision maker only knows the probability distribution of the random variables. This first stage decision is denoted as x. Then the random events happen and full information on them is received, i.e. the realization of the random variables is observed. Then a *second stage decision* $y = y(x, \xi)$ is done. It can be considered as a correction of x because of ξ.
In mathematical terms, the problem with linear objective function and linear constraints reads

$$\begin{aligned} &\min_x \ \mathrm{E}_\xi z(x,\xi) \\ &\text{s.t. } Ax = b,\ x \geq 0 \end{aligned} \tag{2.1}$$

where

$$z(x,\xi) = c^T x + Q(x,\xi) \tag{2.2}$$

and

$$Q(x,\xi) = \min_y \{q(\xi)^T y : W(\xi)y = h(\xi) - T(\xi)x, y \geq 0\}.$$

This formulation will be used many times in this work, usually without repeating the following specification: The first stage decision x is a real vector, $x \in \mathbb{R}^n$, the second stage decision y is a function of x and of (the realization of) ξ, $y = y(x,\xi) \in \mathbb{R}^m$ almost surely. ξ is a random vector defined on some probability space $(\Omega, \mathscr{A}, P)$ with a support $\Xi \subseteq \mathbb{R}^k$. The objective function $z\colon \mathbb{R}^n \times \Xi \to \mathbb{R}$ as defined in (2.2), unless stated otherwise, is always supposed to be continuous and convex[2] in the variable x. The first stage costs vector $c \in \mathbb{R}^n$ is given (non-random), as well as the matrix $A \in \mathbb{R}^l \times \mathbb{R}^n$ and the vector $b \in \mathbb{R}^l$ in the first stage constraints. Instead of the constraint $Ax = b$, $x \geq 0$, we will often write $x \in K$ for a convex polyhedral set $K \subseteq \mathbb{R}^n$. The second stage costs vector q is a function of ξ, so it is a random vector and we suppose that $q(\xi) \in \mathbb{R}^m$ for all $\xi \in \Xi$. The so called *technological matrix* $T = T(\xi)$ depends on ξ as well and for all $\xi \in \Xi$ we suppose that $T(\xi) \in \mathbb{R}^i \times \mathbb{R}^n$. The matrix T provides the relationship between x and y (which reacts to x and ξ). The matrix $W = W(\xi) \in \mathbb{R}^i \times \mathbb{R}^m$ is called a *recourse matrix*. At last, $h = h(\xi)$ is an i-component real random vector. Note that the dimensions are not important, they only have to be such that the multiplication of vectors and matrices is well defined.
The constraints $x \geq 0$, $y \geq 0$ have the sense $x_j \geq 0$ for all $j = 1, \dots, n$, and $y_j \geq 0$ for all $j = 1, \dots, m$, respectively.
The second stage objective function $Q(x,\xi)$ depends on the random variable and so we have to compute with $\mathrm{F}[Q(x,\xi)]$ instead of $Q(x,\xi)$, where F is a probability functional such that the result $\mathrm{F}[Q(x,\xi)]$ does not depend on the realization of the random variable ξ (but it depends on its distribution). The most often and perhaps most natural case of F is the expectation E as in (2.1) but there are many other possibilities.

[2] The function z is not generally convex. Sufficient conditions for convexity of z are given in the next theorem; they are not too restrictive.

We always suppose that the following assumptions are satisfied:
Assumption 2.1 The two-stage linear problem has an optimal solution.
Assumption 2.2 There exists a constant L such that $c^T x + Q(x, \xi) \geq L$ for all feasible x and for all possible realizations of ξ.

To explain the way how the form of the two-stage problem was found, we can imagine that the decision maker had to solve a problem

$$\begin{aligned} \min_x \; & c^T x \\ \text{s.t. } & Ax = b, \; x \geq 0 \end{aligned}$$

at first. When its optimal solution x^* is determined, the decision maker finds out that there is another (latent) constraint of the form

$$T(\xi)x = h(\xi) \text{ a.s.}$$

which he did not consider before. The decision maker can compensate the difference $T(\xi)x - h(\xi)$ with help of the second stage decision y satisfying the condition

$$W(\xi)y = T(\xi)x - h(\xi) \text{ a.s.},$$

and the price for this correction is $q(\xi)^T y$, which is added to the first stage costs $c^T x$. Putting this together and adding the expectation, we obtain the problem (2.1).

The kind of problems we have just introduced is called *recourse problem.*
When the recourse matrix W is non-random, we speak about problem with *fixed recourse.* When additionally the (vector) equation $Wy = u$ has a nonnegative solution y for every $u \in \{T(\xi)x - h(\xi) : \xi \in \Xi, Ax = b, x \geq 0\}$ we speak about *relatively complete recourse*, while the situation when the same equation $Wy = u$ has a nonnegative solution y for all $u \in \mathbb{R}^i$ is called *complete recourse.* The most special situation, the *simple recourse*, is characterised by $W = (I_1 | -I_2)$ and $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, where for $j = 1, 2$ the I_j is an identity matrix.

Nonlinear two-stage programs represent just a generalization of the linear ones. Again, there are decisions to be made in the first and in the second stage, without and with full information on some random events, respectively. The difference is that some of the objective and recourse function or constraints are no more linear.

Theorem 2.3 Consider the problem (2.1) with complete recourse, such that the set $\{x : Ax = b, x \geq 0\}$ is nonempty, expectation of T and h exists, W and q are non-random and W's rank is i (which is number of W's rows), and $\{u : W^T u \leq q\} \neq \emptyset$.
Then the function $z: \; z(x, \xi) = c^T x + \min_y \{q^T y : Wy = h(\xi) - T(\xi)x, y \geq 0\}$ is convex in x for any fixed ξ.
Proof. See [6]. □

2.1.1. Nonanticipativity

There is a very important feature of recourse problems (or, more generally, of multiperiod problems): Any decision must not depend on future realizations of the random variable nor on future decisions. It can depend only on past decisions and past realizations of random variables, which are already known, i.e., the decisions are functions of past realizations and past decisions only, since the decision maker cannot anticipate (presume) future events. In fact, there is $y = y(x, \xi)$ but there is not $x = x(\xi)$. This feature is called *nonanticipativity.*

It is also possible to reformulate the problem with explicit constraints on nonanticipativity. These constraints make the first stage solution independent on random events, i.e., to be the same for all kinds of future environment. In this formulation the problem reads

$$\begin{aligned}
\min_{(x(\xi),y(\xi))} \quad & \mathrm{E}_\xi[c^T x(\xi) + q(\xi)^T y(\xi)] \\
\text{s.t.} \quad & Ax(\xi) = b \text{ a.s.}, \ x(\xi) \geq 0 \text{ a.s.} \\
& W(\xi)y(\xi) = h(\xi) - T(\xi)x(\xi) \text{ a.s.}, \\
& y(\xi) \geq 0 \text{ a.s.}, \\
& \mathrm{E}_\xi[x(\xi)] - x(\xi) = 0 \text{ a.s.}
\end{aligned} \tag{2.3}$$

or

$$\begin{aligned}
\min_{(\tilde{x},x(\xi),y(\xi))} \quad & \mathrm{E}_\xi[c^T x(\xi) + q(\xi)^T y(\xi)] \\
\text{s.t.} \quad & Ax(\xi) = b \text{ a.s.}, \ x(\xi) \geq 0 \text{ a.s.}, \\
& W(\xi)y(\xi) = h(\xi) - T(\xi)x(\xi) \text{ a.s.}, \\
& y(\xi) \geq 0 \text{ a.s.}, \\
& x(\xi) = \tilde{x} \text{ a.s.}, \\
& \tilde{x} \in \mathbb{R}^n.
\end{aligned}$$

2.1.2. Scenario representation of two-stage stochastic problems

Our situation is usually much simpler if we know that random variables included in the problem are of discrete distributions. Then we can represent the future development via scenarios. Each scenario represents one state of world. The resulting environment can be the same for distinct scenarios but they differ in the history leading to the resulting state of world. In mathematical terms, we assume that the support Ξ of ξ is finite, $\Xi = \{\xi^s, \ s = 1, \ldots, S\}$.
Stochastic two-stage problem is then

$$\begin{aligned}
& \min_x \ \left\{c^T x + \mathrm{E}_\xi Q(x, \xi)\right\} \\
& \text{s.t.} \ \ Ax = b, \ x \geq 0,
\end{aligned} \tag{2.4}$$

where $\mathrm{E}_\xi Q(x, \xi) = \sum_{s=1}^{S} P(\xi = \xi^s).Q(x, \xi^s)$ and for all $s = 1, \ldots, S$ it holds

$$\begin{aligned}
Q(x, \xi^s) = \min_{y(\xi^s)} \ & \left\{q(\xi^s)^T y(\xi^s)\right\} \\
\text{s.t. } & W(\xi^s)y(\xi^s) = h(\xi^s) - T(\xi^s)x, \\
& y(\xi^s) \geq 0.
\end{aligned}$$

Here every scenario ξ^s determinates all components of $q, \ h, \ T$ and W, and for every scenario ξ^s is found an optimal second stage solution $y(\xi^s)$, while the first stage solution is same for all scenarios.
In most of real-life situations, we are able to replace all continuous random variables by their reasonable discrete approximations.

2.2. Basic approaches to two-stage stochastic problems

In this section we will deal with linear stochastic problems of the form (2.1). We will define some quantities connected with the recourse problem and with different approaches of solving it. We will also define two basic information value types.

2.1.3. Here-and-now problem

We assume there that z which is given in (2.2) is such that $\mathrm{E}_\xi z(x,\xi)$ exists for all x and also that the minimum of $\mathrm{E}_\xi z(.,\xi)$ is attained for some x. We can also assume here that $c \neq 0$.

When the decision maker solves the stochastic problem (2.1) (the nonanticipative one), he finds a so called *here-and-now* solution. Let us denote this solution as $(x^*, y^*(\xi))$. The optimal value of the objective function is then

$$\begin{aligned} RP = \min_x \ & \mathrm{E}_\xi z(x,\xi) \\ \text{s.t. } & Ax = b, \ x \geq 0 \end{aligned} \tag{2.5}$$

i.e.,

$$RP = c^T x^* + \mathrm{E}_\xi[q(\xi)^T y^*(\xi)].$$

2.1.4. Wait-and-see problem

If the decision maker waits until he has full information on the random events (or if he is a clairvoyant knowing the future in advance) and then he solves a slightly different problem

$$\begin{aligned} WS = \mathrm{E}_\xi \min_x \ & z(x,\xi) \\ \text{s.t. } & Ax = b, x \geq 0, \end{aligned} \tag{2.6}$$

he gets so called *wait-and-see* solution $(\hat{x}(\xi), \hat{y}(\xi))$ for all possible realizations of ξ, and the true expectation of the optimal objective function value

$$WS = \mathrm{E}_\xi[c^T \hat{x}(\xi) + q(\xi)^T \hat{y}(\xi)].$$

It should be already seen, that the wait-and-see solution is not worse than the recourse problem solution, because in the wait-and-see solution even the first stage decision is the best one for each possible realization of ξ, because $x = x(\xi)$.

2.1.5. Expected value problem

Another related problem is called *expected value problem.* The decision maker replaces all random variables by their expected values and solves a simpler deterministic linear program:

$$\begin{aligned} EV = \min_x \ & z(x, \mathrm{E}\xi) = \min_x \left\{ c^T x + \min_y \left\{ q(\mathrm{E}\xi)^T y : W(\mathrm{E}\xi) y = h(\mathrm{E}\xi) - T(\mathrm{E}\xi) x, \ y \geq 0 \right\} \right\} \\ \text{s.t. } & Ax = b, \ x \geq 0. \end{aligned}$$

Let $(\hat{x}(\mathrm{E}\xi), \hat{y}(\mathrm{E}\xi))$ be an optimal solution of this problem. It is unpleasant that the random variable ξ need not attain the value $\mathrm{E}\xi$ at all, e.g. for ξ with an alternative distribution. Then, $\hat{x}(\mathrm{E}\xi)$ may not be useful for the initial problem.

However, we define EEV as expected costs when using $\hat{x}(\mathrm{E}\xi)$ as the first stage decision and a second stage solution which reacts optimally to $\hat{x}(\mathrm{E}\xi)$ and the actual realization of ξ:

$$EEV = \mathrm{E}_\xi z(\hat{x}(\mathrm{E}\xi), \xi) = \mathrm{E}_\xi \left[c^T \hat{x}(\mathrm{E}\xi) + Q(\hat{x}(\mathrm{E}\xi), \xi) \right]. \tag{2.7}$$

Remark 2.4 We consider the vectors q and h and the matrices W and T as functions of ξ. It is clear that in some cases, especially for ξ discrete, these functions may not be defined at the point $\mathrm{E}\xi$. For instance, if ξ has a two-point distribution such that $P(\xi = 0) = P(\xi = 1) = \frac{1}{2}$ and the vector $q = q(\xi)$ is defined as $q(0) = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$, $q(1) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ then we cannot even guess how $q(\mathrm{E}\xi) = q(\frac{1}{2})$ should be defined. In addition, in practice, it is debatable whether we need to deal with $q(\mathrm{E}\xi)$ (costs which are realized with probability 0) or with expected costs $\mathrm{E}q(\xi)$.
To avoid this problem we can consider only random variable ξ such that $\mathrm{E}\xi$ belongs to its support, i.e. $P(\xi = \mathrm{E}\xi) > 0$, and q, h, W and T continuous functions. As a very simple case we can use $q(\xi) = \xi.\tilde{q}$ for a non-random vector $\tilde{q}$ and a scalar random variable ξ.

2.1.6. Expected value of perfect information and value of stochastic solution

Definition 2.5 *Expected value of perfect information* for the problem (2.1) is defined as

$$EVPI = RP - WS, \tag{2.8}$$

where the quantities RP, WS are defined in (2.5) and (2.6), respectively.

The expected value of perfect information can be seen as costs of uncertainty in the problem and also as the value of knowing the future in advance. If the costs for waiting until ξ is observed or the costs for gaining information on the future development of ξ were equal to $EVPI$ (or larger), then there would be no gain from this waiting or from gaining the information.

Another information value compares the results of the expected value approach and the approach of recourse problem.
Definition 2.6 *Value of stochastic solution* for the problem (2.1) is defined as

$$VSS = EEV - RP, \tag{2.9}$$

where the quantities EEV, RP are defined in (2.7) and (2.5), respectively.

This value says how much it is worth doing to solve the stochastic recourse problem rather then the deterministic expected value problem. So VSS is a value of knowing distribution of future outcomes and utilizing it. Nonnegativity of VSS will be shown soon.

Note that solving the wait-and-see problem is unreal, but hypothetically computationally easy, since it is a deterministic problem. The computational demands are low also in the case of the EV and EEV problems, while solving the here-and-now problem is much more demanding.

Let us go back to scenario-based problem (2.4) for a while. A related problem for one particular scenario ξ^s reads

$$\begin{aligned} \min_x z(x, \xi^s) = \min_x & \left\{ c^T x + \min_{y(\xi^s)} \{ q(\xi^s)^T y(\xi^s) : W(\xi^s) y(\xi^s) = h(\xi^s) - T(\xi^s) x, y(\xi^s) \geq 0 \} \right\} \\ \text{s.t. } & x \in K = \{ x \in \mathbb{R}^n : Ax = b, \ x \geq 0 \}. \end{aligned} \tag{2.10}$$

Denote the first stage optimal solution to the problem (2.10) as $\hat{x}(\xi^s)$ and the optimal objective function value as $z(\hat{x}(\xi^s), \xi^s)$. The wait-and-see value is then

$$WS = \mathrm{E}_\xi \left[\min_{x \in K} z(x, \xi) \right] = \sum_{s=1}^{S} P(\xi = \xi^s).z\left(\hat{x}(\xi^s), \xi^s\right).$$

Formulating the EV problem goes in a similar way: Define $\bar{\xi} = \sum_{s=1}^{S} P(\xi = \xi^s).\xi^s$, the first stage expected value problem solution $\hat{x}(\bar{\xi})$ is an optimal solution of the problem $\min_{x \in K} z(x, \bar{\xi})$. This optimal solution is used to compute the EEV value:

$$EEV = \mathrm{E}_\xi z\left(\hat{x}\left(\bar{\xi}\right), \xi\right) = \sum_{s=1}^{S} P(\xi = \xi^s).z\left(\hat{x}\left(\bar{\xi}\right), \xi^s\right).$$

Values of $EVPI$ and VSS are computed as in (2.8) and (2.9), respectively.

2.2. Basic inequalities

In this section some relations between the EV, EEV, WS and RP values will be derived, according to [4] and [2]. We will also show that these inequalities lead to very simple bounds on $EVPI$. Much more about bounding and estimating $EVPI$ and VSS can be found in chapters 2.3.–2.6.

Theorem 2.7 For any two-stage stochastic problem of the form (2.1) and related values WS and EEV defined in (2.6) and (2.7), respectively, it holds

$$WS \leq RP \leq EEV.$$

Proof. The proof can be found in [4]. □

In the following theorem, we use a quite strong premise of non-randomness of q, T and W. This means that the only variable which is allowed to be random is h and so we could separate the randomness to the right hand side of the equation delimitating the second stage feasibility set, and also that the the objective function z is convex in ξ, which is a very strong condition.

Theorem 2.8 For any two stage stochastic problem of the form (2.1) with non-random vector q and non-random matrices T and W it holds

$$EV \leq WS.$$

Proof. The proof can be found in [4]. □

Remark 2.9 In our opinion, the claim $EV \leq WS$ of this theorem is a little unexpected, since the intuition is that the wait-and-see problem gives the "best" (i.e. the lowest) optimal objective function value. In our opinion, this discordance with the intuition is justified by the fact that the assumptions of the theorem define quite a special situation. It is not difficult to find examples that give evidence that the assumption of fixed vector q and fixed matrices T and W which implies convexity of z in ξ cannot be weaken.

Remark 2.10 If q, T and W are deterministic, the only stochastic variable is $h = h(\xi)$ and we can write $h = h(\xi) = \xi$. Hence, the problem reads

$$\begin{aligned} RP = \min_x \ & \left\{ c^T x + \mathrm{E} \min_y \left\{ q^T y : Wy = \xi - Tx,\ y \geq 0 \right\} \right\} \\ \text{s.t. } & Ax = b,\ x \geq 0 \end{aligned}$$

which is equivalent to

$$\begin{aligned} RP = \min_x \ & \left\{ c^T x + \mathrm{E} \min_y q^T y \right\} \\ \text{s.t. } & Ax = b,\ x \geq 0, \\ & y \in K(x, \xi)\ \forall x, \xi, \end{aligned} \tag{2.11}$$

where $K(x, \xi) = \{y : Wy = \xi - Tx,\ y \geq 0\}$.
Assume now, moreover, that W is square and regular, i.e. that the inversion W^{-1} exists. Then we have

$$y = y(x, \xi) = W^{-1}(\xi - Tx) = W^{-1}\xi - W^{-1}Tx \tag{2.12}$$

which is an explicit and unambiguous dependence of y on x and ξ. This means that only one y satisfying (2.12) exists for a given first stage decision x and realization $\xi \in \Xi$. If this y is nonnegative then the set $K(x, \xi)$ is a singleton; otherwise, it is an empty set. We can therefore omit the "$\min_y$" in the objective function of (2.11).
Thus, assuming that W^{-1} exists the problem (2.11) can be written in the form

$$RP = \min_x c^T x + \left[q^T W^{-1} \mathrm{E}\xi - q^T W^{-1} T x \right] \tag{2.13}$$

$$\text{s.t. } Ax = b,\ x \geq 0,$$

$$W^{-1}(\xi - Tx) \geq 0\ \forall \xi \in \Xi. \tag{2.14}$$

The constraint (2.14) implies that $W^{-1}(\mathrm{E}\xi - Tx) \geq 0$ (but the reverse implication does not hold). This further implies that for all x and ξ holds $K(x, \xi) \subseteq K(x, \mathrm{E}\xi)$. Keeping in mind that the form (2.13) of the objective function is the same for the recourse problem and for the expected value problem, we obtain that $RP \geq EV$. When $K(x, \xi) = K(x, \mathrm{E}\xi)$ for all x and for all $\xi \in \Xi$ then $RP = EV$ and so $RP = WS = EV$ according to the theorem 2.8. The other situation $K(x, \xi) \subset K(x, \mathrm{E}\xi)$ for some x would mean that $K(x, \xi)$ is empty for this x and for a certain $\xi \in \Xi$ (since $K(x, \mathrm{E}\xi)$ is a singleton or an empty set) which would mean that there exists a pair x, ξ satisfying $Ax = b, x \geq 0$ and $\xi \in \Xi$ such that there does not exists any feasible second stage solution $y(x, \xi)$ corresponding to this pair. This is impossible to happen in the case of a relatively complete recourse. Therefore, for a two-stage stochastic problem with non-random q, T and W such that W^{-1} exists and with relatively complete recourse it holds $K(x, \xi) = K(x, \mathrm{E}\xi)\ \forall x, \xi$ and so $RP = WS = EV$.

Theorem 2.11 For the problem (2.1) with non-random c, q, T and W holds

$$EV \leq WS \leq RP \leq EEV.$$

and so

$$EVPI = RP - WS \leq RP - EV \tag{2.15}$$

Proof. This follows immediately from theorems 2.7 and 2.8. □

Remark 2.12 We can mention another approach leading to the same upper bound on *EVPI* as in the equation (2.15). Let's have $RP = \min_x \mathrm{E}_\xi z(x,\xi)$ and $WS = \mathrm{E}_\xi \min_x z(x,\xi)$ subject to usual constraints, and suppose that z is convex in x and ξ. Now we define $\psi(\xi) = \min_x z(x,\xi)$ and suppose that this minimum exists for all $\xi \in \Xi$. Then ψ is convex (by Iglehart's lemma) and we have for any point ξ^0 from ψ's domain that

$$\mathrm{E} \min_x z(x,\xi) \geq \min_x z(x,\xi^0) + \nabla\psi(\xi^0)^T(\mathrm{E}\xi - \xi^0)$$

where the symbol ∇ stands for a gradient (if it exists) or a subgradient. Then

$$\begin{aligned} EVPI &= \min_x \mathrm{E}z(x,\xi) - \mathrm{E}\min_x z(x,\xi) \\ &\leq \min_x \mathrm{E}z(x,\xi) - \left(\min_x z(x,\xi^0) + \nabla\psi(\xi^0)^T(\mathrm{E}\xi - \xi^0)\right) \ \forall \xi^0 \in \Xi. \end{aligned} \tag{2.16}$$

This upper bound on *EVPI* can be constructed for all possible ξ^0 and a natural question is, for which ξ^0 this bound is the tightest. As is shown in [2], the best possible bound of the type (2.16) is obtained by choosing $\xi^0 = \mathrm{E}\xi$.
If we set $\xi^0 = \mathrm{E}\xi$ in the equation (2.16), we obtain

$$EVPI \leq \min_x \mathrm{E}z(x,\xi) - \min_x z(x,\mathrm{E}\xi) = RP - EV. \tag{2.17}$$

We now denote $x(\xi^0) = \operatorname{argmin}_x z(x,\xi^0)$. Then for any ξ^0 we have $\min_x \mathrm{E}z(x,\xi) \leq \mathrm{E}z(x(\xi^0),\xi)$ and combining this with the inequality (2.17), we obtain

$$EVPI \leq \mathrm{E}z(x(\xi^0),\xi) - \min_x z(x,\mathrm{E}\xi),$$

which is a bound on *EVPI* that is quite easy to evaluate.
Theorem 2.13 For any stochastic program of the form (2.1) it holds
(a) $EVPI \geq 0$ and $VSS \geq 0$.
(b) If the vectors c, q and the matrices T, W are fixed, then $EVPI \leq EEV - EV$ and $VSS \leq EEV - EV$.
Proof. It is an easy consequence of theorems 2.7 and 2.8. □

Remark 2.14 Part (a) of this theorem ensures us that it is worth doing to ask a clairvoyant about future or postpone making our decisions until realization of ξ is observed, and also to solve the stochastic program rather than its deterministic expected value simplification.

Remark 2.15 Theorem 2.13 allows us to bound *EVPI* and *VSS* by $EEV - EV$ which can be easily computed (without solving the stochastic problem). We can also see that if $EEV = EV$ then $VSS = EVPI = 0$. A sufficient (but quite extreme) condition for this to happen is that the optimal wait-and-see solution $\hat{x}(\xi)$ does not depend on ξ. It means that the optimal solution does not react to the realization of ξ and so it is redundant to solve the recourse problem – solving the problem for one fixed ξ from the support is enough.
No inequalities hold in general between *EVPI* and *VSS*, i.e. it is not true generally that $EVPI \leq VSS$ or vice versa, as could be proven via counterexamples. Some examples illustrating a situation where $EVPI = 0$ and $VSS > 0$, and vice versa, can be found in [4], page 142.

It could also seem that both *VSS* and *EVPI* are the larger, the more "randomness" is in the problem, and the "randomness" is naturally represented by the variance $\operatorname{var}\xi$. Sometimes it is true and sometimes not. For examples, see [4], pages 144–145.

2.3. Bounds on the value of information

We will introduce some interesting bounds on the values of information, as well as some new types of the value of information. They are based on convexity of the objective function z and on the Jensen's inequality. We follow the results presented in [10] and [11] and we add some new results and proofs as well.

2.3.1. Information structures

Consider a stochastic problem

$$\begin{aligned} &\min_x \mathrm{E}_\xi z(x,\xi) \\ &\text{s.t. } x \in K, \end{aligned} \tag{2.18}$$

where $K \subseteq \mathbb{R}^n$ is a convex polyhedral set, $\xi\colon \Omega \to \Xi \subseteq \mathbb{R}$ is a scalar random variable with known probability distribution and a support Ξ. Denote as $F(\xi)$ the cumulative distribution function of ξ. The function $z\colon K \times \Xi \to \mathbb{R}$ is convex in x and ξ and the value $z(x,\xi)$ has the meaning of net costs of the decision maker when he applies his decision x and the random variable is observed to be ξ. Recourse, if any, is included in $z(x,\xi)$.
Let us denote

$$Z_n = \min_{x\in K} \mathrm{E}_\xi z(x,\xi) = RP$$

an optimal objective function value of the recourse problem; the subscript n stands for "no information." Similarly,

$$Z_p = \mathrm{E}_\xi \min_{x\in K} z(x,\xi) = WS$$

will be the optimal objective function value under perfect information, and we define

$$EVPI = Z_n - Z_p.$$

Definition 2.16 η is an *information structure*, if it consists of a discrete set $Y = \{y_i, i \in I\}$ of signal values of a discrete signal random variable y such that when $y_i \in Y$ is observed then $F(\xi)$ changes to $F(\xi|y_i)$, and defining a partition of Ξ into pairly disjoint nonempty subsets $\Xi_i,\ i \in I$ covering Ξ, such that $P(\xi \in \Xi_i | y = y_i) = 1$.
Definition 2.17 Information structure η^2 is *finer* then η^1 if the partition generated by η^2 is finer then that generated by η^1, i.e., for every subset Ξ_i^2 from the partition of Ξ related to η^2 there exists a subset Ξ_j^1 from the partition of Ξ related to η^1 such that $\Xi_i^2 \subseteq \Xi_j^1$.
If an information structure η is available, the stochastic problem (2.18) changes into the stochastic problem

$$Z_\eta = \mathrm{E}_y \min_{x\in K} \mathrm{E}_{\xi|y} z(x,\xi)$$

where

$$\mathrm{E}_{\xi|y} z(x,\xi) = \mathrm{E}_\xi[z(x,\xi)|y] = \int_{-\infty}^{\infty} z(x,\xi)dF(\xi|y)$$

is the conditional expectation of ξ given y, and E_y is expectation with respect to the possible signals of η.

If ξ is discrete and ξ and y (relating to η) are perfectly correlated then $Z_\eta = Z_p$.

Theorem 2.18 $Z_{\eta^2} \leq Z_{\eta^1}$ for every objective function z and every distribution of ξ if and only if η^2 is at least as fine as η^1.

Proof. See [10] for references. □

This theorem ensures us again that $EVPI \geq 0$, since the perfect information structure is finer then the no information structure.

Now we can use the partial information structures to construct some bounds on the expected value of partial information.

Theorem 2.19 (Conditional Jensen's inequality bounds on $EVPI$)
Consider a stochastic problem (2.18). Suppose that z is convex in (x, ξ) on $K \times \Xi$ and an information structure η is available. Then

$$EVPI \leq \min_{y \in Y} \mathrm{E}_\xi z(\bar{x}(y), \xi) - \mathrm{E}_y z(\bar{x}(y), \bar{\xi}(y)),$$

where $\bar{x}(y)$ is an optimal solution to $\min_{x \in K} z(x, \bar{\xi}(y))$ and $\bar{\xi}(y) = \mathrm{E}[\xi|y] = \int_{-\infty}^{\infty} \xi dF(\xi|y)$.

Proof. There is

$$\begin{aligned}
EVPI &= \min_{x \in K} \mathrm{E}_\xi z(x, \xi) - \mathrm{E}_\xi \min_{x \in K} z(x, \xi) = \min_{x \in K} \mathrm{E}_\xi z(x, \xi) - \mathrm{E}_y \mathrm{E}_{\xi|y} \min_{x \in K} z(x, \xi) \\
&\leq \min_{x \in K} \mathrm{E}_\xi z(x, \xi) - \mathrm{E}_y \min_{x \in K} z(x, \bar{\xi}(y)) \text{ by Iglehart's lemma and Jensen's inequality} \\
&= \min_{x \in K} \mathrm{E}_\xi z(x, \xi) - \mathrm{E}_y z(\bar{x}(y), \bar{\xi}(y)) \text{ by definition of } \bar{x}(y) \\
&\leq \min_{y \in Y} \mathrm{E}_\xi z(\bar{x}(y), \xi) - \mathrm{E}_y z(\bar{x}(y), \bar{\xi}(y)) \text{ by feasibility of } \bar{x}(y),\ y \in Y.
\end{aligned}$$

□

Remark 2.20 The convexity of z in (x, ξ) is a very strong condition. The function z as given in (2.2) is convex in x under (quite mild) conditions of the theorem 2.3, but for convexity of z in ξ, non-random q, T and W are necessary, so we deal with a problem with random right hand side only.

2.3.2. Generalized Jensen and Emundson-Madansky bounds

In this paragraph we will partially follow [10] and [11]. We will again deal with the optimization problem (2.18).

Suppose that ξ is a random variable with distribution function $F(\xi)$ and support $\langle a, b\rangle \subseteq \mathbb{R}$, consider a function $f(\xi)$ convex on the closed interval $\langle a, b\rangle$. Then Jensen's inequality gives us $\mathrm{E}f(\xi) \geq f(\mathrm{E}\xi) = f(\bar{\xi})$ and we will denote $f(\bar{\xi}) = J^0$. We can define a function $s_{\langle a,b\rangle}$: $s_{\langle a,b\rangle}(\xi) = \left(\frac{b-\xi}{b-a}\right).f(a) + \left(\frac{\xi-a}{b-a}\right).f(b)$.

Function f is convex and so $f(\xi) \leq s_{\langle a,b\rangle}(\xi)$ for all $\xi \in \langle a, b\rangle$ and

$$\begin{aligned}
J^0 \leq \mathrm{E}f(\xi) = \int_a^b f(\xi) dF(\xi) &\leq \int_a^b s_{\langle a,b\rangle}(\xi) dF(\xi) \\
&= \int_a^b \left[\left(\frac{b-\xi}{b-a}\right).f(a) + \left(\frac{\xi-a}{b-a}\right).f(b)\right] dF(\xi) \\
&= \left(\frac{b-\bar{\xi}}{b-a}\right).f(a) + \left(\frac{\bar{\xi}-a}{b-a}\right).f(b) = s_{\langle a,b\rangle}(\bar{\xi}).
\end{aligned}$$

We denote $s_{\langle a,b\rangle}(\bar{\xi}) = M^0$.

Let us generalize the idea of J^0 and M^0 : For a while we can assume for simplicity that $\alpha_1 = P(\xi \in \langle a, \bar{\xi}\rangle) = P(\xi \in \langle \bar{\xi}, b\rangle) = \alpha_2 = \frac{1}{2}$.

Then we can define $J^1 = \frac{1}{2}f(\bar{\xi}^1) + \frac{1}{2}f(\bar{\xi}^2)$ where we have $\bar{\xi}^1 = (\alpha_1)^{-1}\int_a^{\bar{\xi}} \xi dF(\xi)$ and $\bar{\xi}^2 = (\alpha_2)^{-1}\int_{\bar{\xi}}^b \xi dF(\xi)$. Analogically, we can define M^1 as $\alpha_1 M_1^1 + \alpha_2 M_2^1$ where M_1^1 and M_2^1 are the points found in the same way as M^0 on the intervals $\langle a, \bar{\xi}\rangle$ and $\langle \bar{\xi}, b\rangle$, respectively. The meaning of these definitions is clear from the figure 2.21.

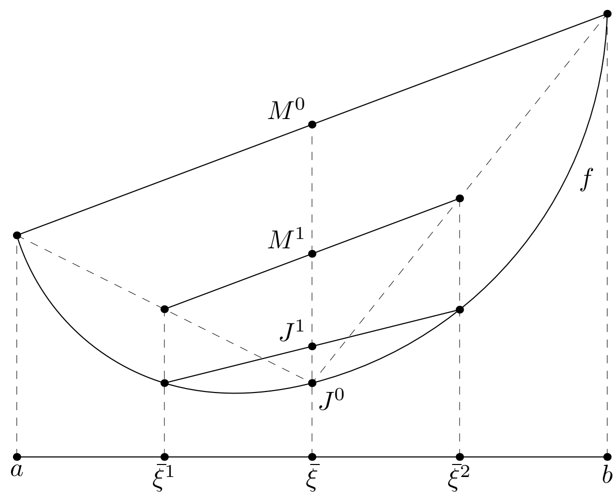

Figure 2.21.

Then

$$\begin{aligned} J^0 = f(\mathrm{E}\xi) \le J^1 &= \frac{1}{2}f(\bar{\xi}^1) + \frac{1}{2}f(\bar{\xi}^2) \le \mathrm{E}f(\xi) \\ &\le \frac{1}{2}\left[\frac{\bar{\xi}-\bar{\xi}^1}{\bar{\xi}-a}.f(a) + \frac{\bar{\xi}^1-a}{\bar{\xi}-a}.f(\bar{\xi})\right] + \frac{1}{2}\left[\frac{b-\bar{\xi}^2}{b-\bar{\xi}}.f(\bar{\xi}) + \frac{\bar{\xi}^2-\bar{\xi}}{b-\bar{\xi}}.f(b)\right] \\ &= \frac{1}{2}s_{\langle a,\bar{\xi}\rangle}(\bar{\xi}^1) + \frac{1}{2}s_{\langle \bar{\xi},b\rangle}(\bar{\xi}^2) = \frac{1}{2}M_1^1 + \frac{1}{2}M_2^1 = M^1 \le s_{\langle a,b\rangle}(\bar{\xi}) = M^0. \end{aligned}$$

Even more general and more precise definitions and notation are introduced in the following theorem.

Theorem 2.22 Suppose that $\langle a, b\rangle$ is an interval subdivided at arbitrary points $d_0, \dots, d_n$, $n \in \mathbb{N}$, where $a = d_0 < d_1 < \dots < d_n = b$. Let $f(\xi)$ be a continuous convex function of a scalar random variable ξ on its support $\langle a, b\rangle$, let F be the distribution function of ξ. Define

$$\begin{aligned} \alpha_i &:= \int_{d_{i-1}}^{d_i} dF(t) > 0,\ \beta_i := \frac{1}{\alpha_i}\int_{d_{i-1}}^{d_i} t dF(t),\ i = 1,\dots,n, \\ \delta_i &:= \alpha_i\left[\frac{\beta_i - d_{i-1}}{d_i - d_{i-1}}\right] + \alpha_{i+1}\left[\frac{d_{i+1}-\beta_{i+1}}{d_{i+1}-d_i}\right],\ i = 0,\dots,n, \\ \alpha_0 &= \alpha_{n+1} = \beta_0 = \beta_{n+1} = d_{-1} := 0. \end{aligned}$$

Let J^n and M^n denote the generalized Jensen and Emundson-Madansky (GJEM) bounds of the n-th order, respectively, defined as:

$$J^n := \sum_{i=1}^n \alpha_i f(\beta_i),\ M^n := \sum_{i=0}^n \delta_i f(d_i),\ n = 1, 2, \dots$$

Then

(a) $$J^n \le \mathrm{E}f(\xi) \le M^n,\ n = 1, 2, \dots.$$

(b) If the partition of $\langle a, b\rangle$ corresponding to $k+1$ is at least as fine as that corresponding to k for $k = 1, \dots, n-1$, then

$$J^1 \le J^2 \le \dots \le J^n \le \mathrm{E}f(\xi) \le M^n \le \dots \le M^1.$$

(c) If in addition each subinterval becomes arbitrarily small as $n \to \infty$, then

$$\lim_{n\to\infty} J^n = \mathrm{E}f(\xi) = \lim_{n\to\infty} M^n.$$

Proof. See [10] and a reference given there. □

These bounds can naturally be used to compute bounds on $EVPI$.

Theorem 2.23 Suppose that $z(x, \xi)$ is a continuous function convex in (x, ξ) on the convex set $K \times \Xi$ where $\Xi = \langle a, b\rangle$ is the support of ξ and K is a convex and compact set. For $n \in \mathbb{N}$ define

$$L_l(z) = \sum_{i=1}^{l} \alpha_i z(x, \beta_i), \quad U_l(z) = \sum_{i=0}^{l} \delta_i z(x, d_i),\ l = 1, \ldots, n,$$

$$J^l = L_l(\min z) = \sum_{i=1}^{l} \alpha_i \min_{x\in K} z(x, \beta_i), \quad M^l = U_l(\min z) = \sum_{i=0}^{l} \delta_i \min_{x\in K} z(x, d_i),\ l = 1, \ldots, n,$$

where $d_i, \alpha_i, \beta_i, \delta_i$ for $i = 0, \ldots, n+1$ and d_{-1} are defined as in theorem 2.22. Then it holds

(a)
$$\min_{x\in K} L_l(z) \le \min_{x\in K} \mathrm{E}_\xi z(x, \xi) = Z_n \le \min_{x\in K} U_l(z),$$

(b)
$$L_l(\min z) = J^l \le \mathrm{E}_\xi \min_{x\in K} z(x, \xi) = Z_p \le M^l = U_l(\min z),$$

(c)
$$\max\{0; \min_{x\in K} L_l(z) - U_l(\min z)\} \le EVPI \le \min_{x\in K} U_l(z) - L_l(\min z).$$

(d) If the partition corresponding to $l+1$ is finer then that corresponding to l, then the bounds for $l+1$ are at least as sharp as those for l.

(e) If each subinterval becomes arbitrarily small as $l \to \infty$, then

$$\lim_{l\to\infty} [\min_{x\in K} L_l(z) - U_l(\min z)] = EVPI = \lim_{l\to\infty} [\min_{x\in K} U_l(z) - L_l(\min z)].$$

Proof. Parts (a) and (b) are easy consequences of the theorem 2.22. Part (c) is a direct consequence of (a) and (b), (d) follows from monotonicity of J^l and M^l in l, (e) follows from (a)–(d) and the theorem 2.22 (c). □

This theorem can be generalized for the case of vector random variable ξ with independent components; this generalization can be found in [10].

2.3.3. Convex negative-utility function

In general, we can incorporate a *negative-utility* or, say, badness function $b\colon \mathbb{R} \to \mathbb{R}$ which expresses our opinion to varying costs $z(x, \xi)$. The most usual situation is that b is nondecreasing (large costs are worse than low ones) and convex (increase of small costs by 1 unit is felt worse than same increase of large costs). If the negative-utility function b is linear then it can be omitted.

Everything above holds true for a linear *negative-utility function* which can be omitted. Now we will suppose that the negative-utility function is a convex nondecreasing function of costs which expresses the subjective decision maker's feeling of the overall costs and can therefore be considered as a counterpart of the well known utility function which is used in maximization problems. It is natural that higher costs are felt as worse than lower costs and that the decision maker minimizes his negative-utility ("badness") and therefore he minimizes costs.

The decision maker tries to minimize his negative utility function $b\colon \mathbb{R} \to \mathbb{R}$, where $b(z(x,\xi))$ is the badness of the costs $z(x,\xi)$ and z is a continuous and convex function. It is clear that the function b brings nonlinearity into the problem, but in addition we can consider a general convex cost function z now (not only the function z defined in (2.2)).

The here-and-now problem now reads

$$\begin{aligned} &\min_x \; \mathrm{E}_\xi b[z(x,\xi)] \\ &\text{s.t. } x \in K, \end{aligned} \tag{2.19}$$

where K is a convex set. The following definition gives a relevant counterpart of the expected value of perfect information.

Definition 2.24 *Value of information* $\tilde{V}$ relating to the problem (2.19) is a solution of the equation

$$\mathrm{E}_\xi \min_{x\in K} b[z(x,\xi) + \tilde{V}] = \min_{x\in K} \mathrm{E}_\xi b[z(x,\xi)]. \tag{2.20}$$

Remark 2.25 If b is a linear function given as $b(t) = st + r$ then

$$\mathrm{E}_\xi \min_{x\in K} b[z(x,\xi) + \tilde{V}] = \mathrm{E}_\xi \min_{x\in K} [s.z(x,\xi)] + s.\tilde{V} + r$$

and it has to be equal to

$$\min_{x\in K} \mathrm{E}_\xi [s.z(x,\xi) + r] = \min_{x\in K} \mathrm{E}_\xi [s.z(x,\xi)] + r$$

and so $\tilde{V} = EVPI$.

We will derive some bounds on the value of information $\tilde{V}$ now.

Theorem 2.26 Suppose that b is convex strictly increasing on $\mathbb{R}$, z is convex in (x,ξ) on a convex set $K \times \Xi$. Then

$$\tilde{V} \le b^{-1}\left[\mathrm{E}_\xi b[z(\bar{x},\xi)]\right] - z(\bar{x},\bar{\xi}),$$

where $\bar{x}$ is an optimal solution to $\min\limits_{x\in K} z(x,\bar{\xi})$, $\bar{\xi} = \mathrm{E}\xi$, and $\tilde{V}$ is defined in (2.20).

Proof. There is

$$\min_{x\in K} \mathrm{E}_\xi b[z(x,\xi)] \le \mathrm{E}_\xi b[z(\bar{x},\xi)] \tag{2.21}$$

since $\bar{x} \in K$, and

$$\begin{aligned} \mathrm{E}_\xi \min_{x\in K} b[z(x,\xi) + \tilde{V}] &= \mathrm{E}_\xi h(\xi) \text{ where } h(\xi) = \min_{x\in K} b[z(x,\xi) + \tilde{V}] \\ &\ge h(\mathrm{E}\xi) \text{ by Jensen's inequality; } h \text{ is convex by Iglehart's lemma} \\ &= \min_{x\in K} b[z(x,\bar{\xi}) + \tilde{V}] = b[\min_{x\in K} z(x,\bar{\xi}) + \tilde{V}] \text{ since } b \text{ is strictly increasing} \\ &= b[z(\bar{x},\bar{\xi}) + \tilde{V}] \text{ by definition of } \bar{x}. \end{aligned} \tag{2.22}$$

Inequalities (2.21) and (2.22) give together:

$$b[z(\bar{x},\bar{\xi})+\tilde{V}] \le \mathrm{E}_{\xi} b[z(\bar{x},\xi)]$$

which implies that

$$\tilde{V} \le b^{-1}\left[\mathrm{E}_{\xi} b\left[z(\bar{x},\xi)\right]\right] - z(\bar{x},\bar{\xi}).$$

□

Theorem 2.27 Suppose that b is convex strictly increasing on $\mathbb{R}$, z is convex in (x,ξ) on a convex set $K \times \Xi$, an information structure η (with relating signal variable y) is available. Then

$$\tilde{V} \le b^{-1}\left[\min_{y\in Y} \mathrm{E}_{\xi} b[z(\bar{x}(y),\xi)]\right] - \mathrm{E}_y \min_{x\in K} z(x,\bar{\xi}(y)),$$

where $\bar{x}(y)$ is an optimal solution to $\min_{x\in K} z(x,\bar{\xi}(y))$ and $\bar{\xi}(y) = \mathrm{E}[\xi|y] = \int_{-\infty}^{\infty} \xi dF(\xi|y)$ is a conditional expectation of ξ given y and $\tilde{V}$ is defined as a solution of (2.20).

Proof. The idea of the proof is the same as the previous one. See [10] for details. □

As is shown in [10], generalization of all of these results for a vector random variable $\xi^r = (\xi_1,\ldots,\xi_r)$ with independent components is possible. It can be used to derive an upper bound on $\tilde{V}$ which is similar to that introduced in the theorem 2.26 with $\mathrm{E}_{\xi} b[z(\bar{x},\xi)]$ and $z(\bar{x},\bar{\xi})$ replaced by their iterative upper and lower bounds, respectively, as introduced in theorem 2.23. As is mentioned in the cited articles, this method does not lead to a similar lower bound on $\tilde{V}$ (unlike in a problem with a linear or negative-utility function). We now introduce a new lower bound on $\tilde{V}$ for a problem with a convex negative-utility function.

To get a lower bound on $\tilde{V}$, we can use convexity of the negative-utility function b. According to the equation (2.20) and thanks to convexity, for any pair $p,q \in (0,1)$ such that $p+q=1$, we can write

$$\begin{aligned}\min_{x\in K} \mathrm{E}_{\xi^r} b[z(x,\xi^r)] &= \mathrm{E}_{\xi^r} \min_{x\in K} b[z(x,\xi^r)+\tilde{V}] \\ &= \mathrm{E}_{\xi^r} b\left[p.\min_{x\in K}\frac{1}{p} z(x,\xi^r) + q.\frac{1}{q}\tilde{V}\right] \le p.\mathrm{E}_{\xi^r} b\left[\min_{x\in K}\frac{1}{p} z(x,\xi^r)\right] + q.b\left[\frac{1}{q}\tilde{V}\right]\end{aligned}$$

and further operations imply that

$$\begin{aligned} q.b\left[\frac{1}{q}\tilde{V}\right] &\ge \min_{x\in K}\mathrm{E}_{\xi^r} b[z(x,\xi^r)] - p.\mathrm{E}_{\xi^r} b\left[\min_{x\in K}\frac{1}{p} z(x,\xi^r)\right], \\ \tilde{V} &\ge q.b^{-1}\left[\frac{1}{q}\min_{x\in K}\mathrm{E}_{\xi^r} b[z(x,\xi^r)] - \frac{p}{q}\mathrm{E}_{\xi^r} b\left[\min_{x\in K}\frac{1}{p} z(x,\xi^r)\right]\right] \\ &= (1-p).b^{-1}\left[\frac{1}{1-p}\min_{x\in K}\mathrm{E}_{\xi^r} b[z(x,\xi^r)] - \frac{p}{1-p}\mathrm{E}_{\xi^r} b\left[\frac{1}{p}\min_{x\in K} z(x,\xi^r)\right]\right].\end{aligned}$$

The function b is supposed to be strictly increasing and so b^{-1} is strictly increasing as well. Hence, we can substitute $\min_{x\in K}\mathrm{E}_{\xi^r} b[z(x,\xi^r)]$ with any its lower bound $\tilde{L}$ and

$\mathrm{E}_{\xi^r} b\left[\frac{1}{p}\min_{x\in K} z(x,\xi^r)\right]$ with any its upper bound $\tilde{U}(p)$ and the last inequality will hold a fortiori. Hence, it holds

$$\tilde{V} \geq (1-p).b^{-1}\left[\frac{1}{1-p}\tilde{L} - \frac{p}{1-p}\tilde{U}(p)\right] \ \forall p \in (0,1)$$

and so

$$\tilde{V} \geq \sup_{p\in(0,1)} (1-p).b^{-1}\left[\frac{1}{1-p}\tilde{L} - \frac{p}{1-p}\tilde{U}(p)\right].$$

Lower bound of the same type could naturally be used in the case of scalar random variable ξ. Unfortunately, these bounds are quite sophisticated and we can guess that they are generally not too sharp. They are the sharper, the less the function b differs from a linear function.

Much simpler lower bound on V can also be found under an assumption that b is subaditive, i.e., $b(x+y) \leq b(x) + b(y)$ for all $x, \ y$. Then we have

$$\min_{x\in K} \mathrm{E}_{\xi^r} b[z(x,\xi^r)] = \mathrm{E}_{\xi^r} b\left[\min_{x\in K} z(x,\xi^r) + \tilde{V}\right] \leq \mathrm{E}_{\xi^r} b\left[\min_{x\in K} z(x,\xi^r)\right] + b[\tilde{V}]$$

which implies that

$$\begin{aligned}\tilde{V} &\geq b^{-1}\left[\min_{x\in K} \mathrm{E}_{\xi^r} b[z(x,\xi^r)] - \mathrm{E}_{\xi^r} b\left[\min_{x\in K} z(x,\xi^r)\right]\right] \\ &= b^{-1}\left[\min_{x\in K} \mathrm{E}_{\xi^r} b[z(x,\xi^r)] - \mathrm{E}_{\xi^r} \min_{x\in K} b[z(x,\xi^r)]\right] \\ &= b^{-1}[RP - WS],\end{aligned} \tag{2.23}$$

where the RP and WS are now related to the problems with a compound objective function $b(z)$.

Again, $\min_{x\in K} \mathrm{E}_{\xi^r} b[z(x,\xi^r)]$ and $\mathrm{E}_{\xi^r} b\left[\min_{x\in K} z(x,\xi^r)\right]$ in the expression (2.23) can be substituted with their lower and upper bounds, respectively.

Under special distributional assumptions, another lower bound on $\tilde{V}$ for the case of a vector random variable ξ^r was derived in [11].

Theorem 2.28 Suppose that

(1) b is convex strictly increasing on $\mathbb{R}$,

(2) $x^* = x^*(\xi^r)$ solves $\min_{x\in K} z(x,\xi^r)$ and $\hat{x}$ solves $\min_{x\in K} \mathrm{E}^r_\xi b[z(x,\xi^r)]$,

(3) $z(x^*,\xi^r), \ z(\hat{x},\xi^r)$ have distributions from the same family with two parameters that are independent functions of mean and variance, i.e. if $z(x^*,\xi^r) \sim G(y;a_1,b_1)$ and $z(\hat{x},\xi^r) \sim H(z;a_2,b_2)$ then $\frac{y-a_1}{\sqrt{b_1}} = \frac{z-a_2}{\sqrt{b_2}}$ implies that $G(y) = H(z)$ (where a_1, a_2 are finite and b_1, b_2 are finite and positive). Then

(a) $$\tilde{V} \geq \mathrm{E}_{\xi^r}[z(\hat{x},\xi^r) - z(x^*,\xi^r)] \Leftrightarrow \mathrm{var}\, z(x^*,\xi^r) \geq \mathrm{var}\, z(\hat{x},\xi^r).$$

(b) For any function z convex in (x,ξ^r) and any b convex and strictly increasing negative-utility function, there is

$$\mathrm{var}\, z(x^*,\xi^r) \geq \mathrm{var}\, z(\hat{x},\xi^r) \Rightarrow \tilde{V} \geq EVPI, \tag{2.24}$$

where $\tilde{V}$ is the value of perfect information for b convex negative-utility function which is defined a solution of (2.20), and $EVPI$ is the expected value of perfect information for the same two-stage stochastic problem without the negative-utility function b.

Proof.

(a) By definition, there is $z(x^*, \xi^r) = \min_{x \in K} z(x, \xi^r)$ and $\mathrm{E}_{\xi^r} b[z(\hat{x}, \xi^r)] = \min_{x \in K} \mathrm{E}_{\xi^r} b[z(x, \xi^r)]$. The function b is strictly increasing and so

$$\mathrm{E}_{\xi^r} \min_{x \in K} b[z(x, \xi^r) + \tilde{V}] = \mathrm{E}_{\xi^r} b\left[\min_{x \in K} z(x, \xi^r) + \tilde{V}\right].$$

The value of information $\tilde{V}$ is defined by (2.20). This is equivalent to

$$\begin{aligned}\mathrm{E}_{\xi^r} b\left[\min_{x \in K} z(x, \xi^r) + \tilde{V}\right] &= \mathrm{E}_{\xi^r} b\left[z(x^*, \xi^r) + \tilde{V}\right] = \mathrm{E}_{\xi^r} b\left[z(\hat{x}, \xi^r)\right] \\ &= \mathrm{E}_{\xi^r} b\left[z(x^*, \xi^r) - \{z(x^*, \xi^r) - z(\hat{x}, \xi^r)\}\right].\end{aligned}$$

Let us denote $g(\xi^r) = z(x^*, \xi^r) - \mathrm{E}_{\xi^r}\{z(x^*, \xi^r) - z(\hat{x}, \xi^r)\}$. Then $g(\xi^r)$ and $z(\hat{x}, \xi^r)$ have distributions belonging to the same family described by two parameters that are independent functions of mean and variance with equal means. Then according to [8]

$$\mathrm{E}_{\xi^r} b[g(\xi^r)] \le \mathrm{E}_{\xi^r} b[z(\hat{x}, \xi^r)] \Leftrightarrow \operatorname{var} g(\xi^r) \ge \operatorname{var} z(\hat{x}, \xi^r)$$

for any nondecreasing function b. Then

$$\mathrm{E}_{\xi^r} b[z(x^*, \xi^r) - \mathrm{E}_{\xi^r}\{z(x^*, \xi^r) - z(\hat{x}, \xi^r)\}] \le \mathrm{E}_{\xi^r} b[z(\hat{x}, \xi^r)] \Leftrightarrow \operatorname{var} g(\xi^r) \ge \operatorname{var} z(\hat{x}, \xi^r)$$

and $\mathrm{E}_{\xi^r} b[z(\hat{x}, \xi^r)] = \mathrm{E}_{\xi^r} b[z(x^*, \xi^r) + \tilde{V}]$. Hence,

$$-\mathrm{E}_{\xi^r}\{z(x^*, \xi^r) - z(\hat{x}, \xi^r)\} \le \tilde{V} \Leftrightarrow \operatorname{var} g(\xi^r) \ge \operatorname{var} z(\hat{x}, \xi^r),$$

that is

$$\mathrm{E}_{\xi^r}\{z(\hat{x}, \xi^r) - z(x^*, \xi^r)\} \le \tilde{V} \Leftrightarrow \operatorname{var} g(\xi^r) \ge \operatorname{var} z(\hat{x}, \xi^r).$$

It is clear that $\operatorname{var} g(\xi^r) = \operatorname{var} z(x^*, \xi^r)$ and so finally

$$\mathrm{E}_{\xi^r}\{z(\hat{x}, \xi^r) - z(x^*, \xi^r)\} \le \tilde{V} \Leftrightarrow \operatorname{var} z(x^*, \xi^r) \ge \operatorname{var} z(\hat{x}, \xi^r).$$

(b) We know that $\mathrm{E}_{\xi^r} z(\hat{x}, \xi^r) \ge \min_{x \in K} \mathrm{E}_{\xi^r} z(x, \xi^r)$. Hence,

$$\begin{aligned}\mathrm{E}_{\xi^r}\{z(\hat{x}, \xi^r) - z(x^*, \xi^r)\} &= \mathrm{E}_{\xi^r} z(\hat{x}, \xi^r) - \mathrm{E}_{\xi^r} \min_{x \in K} z(x, \xi^r) \\ &\ge \min_{x \in K} \mathrm{E}_{\xi^r} z(x, \xi^r) - \mathrm{E}_{\xi^r} \min_{x \in K} z(x, \xi^r) \qquad (2.25) \\ &= RP - WS = EVPI.\end{aligned}$$

Then thanks to (a) it holds

$$\operatorname{var} z(x^*, \xi^r) \ge \operatorname{var} z(\hat{x}, \xi^r) \Rightarrow \tilde{V} \ge \mathrm{E}_{\xi^r}\{z(\hat{x}, \xi^r) - z(x^*, \xi^r)\}.$$

The last expression $\mathrm{E}_{\xi^r}\{z(\hat{x}, \xi^r) - z(x^*, \xi^r)\}$ is greater or equal to $EVPI$ according to (2.25).

□

Remark 2.29 The authors of [11] stated that an equivalence holds in (2.24) but they do not give a correct proof of the second implication

$$\tilde{V} \geq EVPI \Rightarrow \operatorname{var} z(x^*, \xi^r) \geq \operatorname{var} z(\hat{x}, \xi^r).$$

Remark 2.30 The conditions that are required to hold for the family of distributions are not too strange. They come from problems of portfolio mean-variance optimization. The described family of distributions includes normal (not lognormal), uniform and two-point equally-likely distributions.

2.3.4. Partial information

Linear negative-utility function

We have already introduced $Z_\eta = \mathrm{E}_y \min_{x\in K} \mathrm{E}_{\xi|y} z(x,\xi)$ where z is defined in (2.2). *Expected value of partial information* (given by the information structure) η is defined as

$$V_\eta = Z_n - Z_\eta = \min_{x\in K} \mathrm{E}_\xi z(x,\xi) - \mathrm{E}_y \min_{x\in K} \mathrm{E}_{\xi|y} z(x,\xi). \tag{2.26}$$

If η^1, η^2 are two information structures then

$$V_{\eta^2} - V_{\eta^1} = (Z_n - Z_{\eta^2}) - (Z_n - Z_{\eta^1}) = Z_{\eta^1} - Z_{\eta^2} = \mathrm{E}_{y^1} \min_{x\in K} \mathrm{E}_{\xi|y^1} z(x,\xi) - \mathrm{E}_{y^2} \min_{x\in K} \mathrm{E}_{\xi|y^2} z(x,\xi).$$

At first, there are two quite easy but quite important inequalities:

Theorem 2.31

(a) There is $V_\eta \geq 0$ for any information structure η.

(b) If η^1, η^2 are two information structures, η^1 is at least as fine as η^2, then $V_{\eta^1} \geq V_{\eta^2}$.

Proof.

(a) $V_\eta = Z_n - Z_\eta \geq 0$ according to theorem 2.18 since η is at least as fine as the non-information structure (denoted by the subscript n).

(b) $V_{\eta^1} - V_{\eta^2} = Z_{\eta^2} - Z_{\eta^1} \geq 0$ because $Z_{\eta^2} \geq Z_{\eta^1}$. □

We will develop some bounds on $V_{\eta^2} - V_{\eta^1}$, which is a value of adding, increasing information or a value of better partial information.

Theorem 2.32 Suppose that z is a convex function of ξ for all fixed $x \in K$. Then

$$\mathrm{E}_{y^1}\left[\mathrm{E}_{\xi|y^1} z(\bar{x}(y^1),\xi)\right] - \mathrm{E}_{y^2} z(\bar{x}(y^2), \bar{\xi}(y^2)) \geq V_{\eta^2} - V_{\eta^1} = Z_{\eta^1} - Z_{\eta^2}$$
$$\geq \mathrm{E}_{y^1} z(\bar{x}(y^1), \bar{\xi}(y^1)) - \mathrm{E}_{y^2}\left[\mathrm{E}_{\xi|y^2} z(\bar{x}(y^2),\xi)\right],$$

where $\bar{x}(y^i)$ is an optimal solution to $\min_{x\in K} z(x, \bar{\xi}(y^i))$ and $\bar{\xi}(y^i) = \mathrm{E}[\xi|y^i] = \int_{-\infty}^{\infty} \xi dF(\xi|y^i)$, for $i = 1, 2$.

Proof. For $i = 1, 2$ holds true:

$$Z_{\eta^i} = \mathrm{E}_{y^i} \min_{x\in K} \mathrm{E}_{\xi|y^i} z(x,\xi) \geq \mathrm{E}_{y^i} \min_{x\in K} z(x, \bar{\xi}(y^i)) = \mathrm{E}_{y^i} z(\bar{x}(y^i), \bar{\xi}(y^i))$$

by conditional Jensen's inequality applied to $f(\xi) = z(x,\xi)$ for fixed x; we assume that this function is convex for all $x \in K$.

Also,

$$Z_{\eta^i} = \mathrm{E}_{y^i}\left[\min_{x\in K} \mathrm{E}_{\xi|y^i} z(x,\xi)\right] \leq \mathrm{E}_{y^i}\left[\mathrm{E}_{\xi|y^i} z(\bar{x}(y^i),\xi)\right]$$

since $\bar{x}(y^i) \in K$. Now we can use that

$$upper(Z_{\eta^1}) - lower(Z_{\eta^2}) \geq Z_{\eta^1} - Z_{\eta^2} = V_{\eta^2} - V_{\eta^1} \geq lower(Z_{\eta^1}) - upper(Z_{\eta^2})$$

where $upper(Z_{\eta^i})$, $lower(Z_{\eta^i})$ is any upper, lower bound on Z_{η^i}, respectively. □

Convex negative-utility function

Having used a convex strictly increasing negative-utility function b, *value of partial information* (given by an information structure) η is $\tilde{V}_\eta$ which is defined as a solution of the equation

$$\mathrm{E}_y \min_{x \in K} \mathrm{E}_{\xi|y} b[z(x,\xi) + \tilde{V}_\eta] = \min_{x \in K} \mathrm{E}_\xi b[z(x,\xi)]. \tag{2.27}$$

(We have added the tilde to distinguish the value of partial information for a linear and convex negative-utility function.)

Again, there are two quite important inequalities:

Theorem 2.33 Suppose that z is a convex function of ξ for all fixed $x \in K$, and b is convex and strictly increasing. Then

(a) for any information structure η, there is $\tilde{V}_\eta \geq 0$.

(b) Let η^1 and η^2 be two information structures. If η^1 is at least as fine as η^2, then $\tilde{V}_{\eta^1} \geq \tilde{V}_{\eta^2}$.

Proof.

(a) For any convex function g and a random variable y, we have

$$\mathrm{E}_y \min_{x \in K} g(x,y) \leq \min_{x \in K} \mathrm{E}_y g(x,y).$$

Now let us have $g(x,y) = \mathrm{E}_{\xi|y} b[z(x,\xi)]$. We then obtain that

$$\mathrm{E}_y \min_{x \in K} \left[\mathrm{E}_{\xi|y} b[z(x,\xi)]\right] \leq \min_{x \in K} \mathrm{E}_y \left[\mathrm{E}_{\xi|y} b[z(x,\xi)]\right] = \min_{x \in K} \mathrm{E}_\xi b[z(x,\xi)].$$

The function b is strictly increasing and $\tilde{V}_\eta$ is a solution of the equation (2.27) and so $\tilde{V}_\eta \geq 0$.

(b) For $i = 1,2$, let us denote as y^i the signal random variable relative to the information structure η^i, and $\tilde{z} = b[z]$. For η^1 at least as fine as η^2 we know that

$$\tilde{Z}_{\eta^1} := \mathrm{E}_{y^1} \min_{x \in K} \mathrm{E}_{\xi|y^1} b[z(x,\xi)] \leq \mathrm{E}_{y^2} \min_{x \in K} \mathrm{E}_{\xi|y^2} b[z(x,\xi)] =: \tilde{Z}_{\eta^2}$$

and from definition of $\tilde{V}_{\eta^i}$, $i = 1,2$, it holds

$$\mathrm{E}_{y^1} \min_{x \in K} \mathrm{E}_{\xi|y^1} b[z(x,\xi) + \tilde{V}_{\eta^1}] = \min_{x \in K} \mathrm{E}_\xi b[z(x,\xi)] = \mathrm{E}_{y^2} \min_{x \in K} \mathrm{E}_{\xi|y^2} b[z(x,\xi) + \tilde{V}_{\eta^2}]. \tag{2.28}$$

Let us suppose for contradiction, that $\tilde{V}_{\eta^1} < \tilde{V}_{\eta^2}$, without loss of generality $\tilde{V}_{\eta^1} = 0$ and $\tilde{V}_{\eta^2} > 0$. Then

$$\begin{aligned} \mathrm{E}_{y^1} \min_{x \in K} \mathrm{E}_{\xi|y^1} b[z(x,\xi) + \tilde{V}_{\eta^1}] &= \mathrm{E}_{y^1} \min_{x \in K} \mathrm{E}_{\xi|y^1} b[z(x,\xi)] \\ &\leq \mathrm{E}_{y^2} \min_{x \in K} \mathrm{E}_{\xi|y^2} b[z(x,\xi)] < \mathrm{E}_{y^2} \min_{x \in K} \mathrm{E}_{\xi|y^2} b[z(x,\xi) + \tilde{V}_{\eta^2}], \end{aligned}$$

since b is strictly increasing. So we have obtained a contradiction with (2.28). □

Remark 2.34 It is clear that for any η is $\tilde{V}_\eta \leq \tilde{V} = \tilde{V}_p$.

2.4. Pairs subproblem

Inspired by the approach introduced in [3], we will deal with pairs and groups subproblems and true expectations of some objective function values related to them. We will also define a modification of *VSS* and finally we will derive new bounds on *VSS* which are very interesting and quite computationally demanding.

Throughout this chapter, we will consider a slightly simplified version of a two stage stochastic program with random right hand side only. The problem reads

$$\begin{aligned}&\min_x \mathrm{E}_\xi z(x,\xi) = \min_x \left\{ c^T x + \mathrm{E}_\xi \min_{y(\xi)} \{ q^T y(\xi) : W y(\xi) = \xi - Tx, y(\xi) \geq 0 \} \right\} \\ &\text{s.t.} \quad Ax = b,\ x \geq 0\end{aligned} \tag{2.29}$$

under assumptions that $\xi = h(\xi)$ is the only random variable, support of ξ is a set Ξ which is finite with K possible realizations of ξ, i.e. $\Xi = \{\xi^1, \ldots, \xi^K\}$. The scenarios $\xi^1, \ldots, \xi^K$ come with probabilities $p^1, \ldots, p^K$, respectively, and $\sum_{k=1}^K p^k = 1$.

Consider a *reference scenario* ξ^u (e.g. $\xi^u = \bar{\xi}$ or ξ^u is the worst-case scenario), denote $p^u = P(\xi = \xi^u)$; it is possible that $\xi^u \notin \Xi$ and then $p^u = 0$. We formulate the *pairs subproblem of* ξ^u *and* $\xi^k \in \Xi$ as

$$\begin{aligned}\min_{(x,y(\xi^u),y(\xi^k))} z^P(x,\xi^u,\xi^k) = &\min_{(x,y(\xi^u),y(\xi^k))} \left\{ c^T x + p^u . q^T y(\xi^u) + (1-p^u) . q^T y(\xi^k) \right\} \\ \text{s.t.} \quad & Ax = b,\ x \geq 0, \\ & W y(\xi^u) = \xi^u - Tx, \\ & W y(\xi^k) = \xi^k - Tx, \\ & y \geq 0.\end{aligned} \tag{2.30}$$

Optimal solution of this problem is denoted as $(\hat{x}^{u,k}, \hat{y}(\xi^u), \hat{y}(\xi^k))$ and the optimal objective function value is $z^P(\hat{x}^{u,k}, \hat{y}^(\xi^u), \hat{y}(\xi^k))$.

There are two special cases:

(1) If we have $\xi^k = \xi^u$, then the objective function is $z^P(x,\xi^u,\xi^u) = z(x,\xi^u)$ and we have a deterministic problem for the reference scenario.

(2) If $\xi^u \notin \Xi$ then $p^u = 0$ and we solve a deterministic problem for the scenario ξ^k, since for $p^u = 0$ is $z^P(x,\xi^u,\xi^k) = z(x,\xi^k)$.

Define *expectation of optimized pairs subproblems* with fixed reference scenario ξ^u (which will be denoted as $EOPS^u$) as

$$EOPS^u = \frac{1}{1-p^u} \sum_{\substack{k=1 \\ \xi^k \neq \xi^u}}^{K} p^k . \min_{(x,y(\xi^u),y(\xi^k))} z^P(x,\xi^u,\xi^k)$$

where for every k we minimize the function z^P under the same constraints as in (2.30).

Note that for $\xi^u \notin \Xi$ there is

$$EOPS^u = \frac{1}{1-0} \cdot \sum_{\substack{k=1 \\ \xi^k \neq \xi^u}}^{K} p^k . \min z(x, \xi^k) = \sum_{\xi^k \in \Xi} p^k . \min z(x, \xi^k) = \mathrm{E}_\xi \min z(x, \xi) = WS.$$

Theorem 2.35 For the two stage stochastic program of the form (2.29) and for any reference scenario ξ^u it holds

$$WS \leq EOPS^u \leq RP.$$

Proof. The proof can be found in [3]. □

Let us generalize the definition of the value of stochastic solution (VSS) for a while. Let $\bar{x}^u$ be an optimal solution of the deterministic problem for the reference scenario ξ^u

$$\min_x z(x, \xi^u) = \min_x \left\{ c^\top x + \min_{y(\xi^u)} \{ q^\top y(\xi^u) : W y(\xi^u) = \xi^u - Tx, y(\xi^u) \geq 0 \} \right\}$$
$$\text{s.t. } Ax = b, \; x \geq 0.$$

Define $EVRS^u = \mathrm{E}_\xi z(\bar{x}^u, \xi)$ and

$$VSS^u = EVRS^u - RP. \tag{2.31}$$

If there is $\xi^u = \bar{\xi} = \mathrm{E}\xi$, then $EVRS^u = EEV$ and nothing has changed. The nonnegativity of VSS is still kept: If $\bar{x}^u$ is a first stage feasible solution of the here-and-now problem then $EVRS^u \geq RP$, and if $\bar{x}^u$ is infeasible then $EVRS^u = +\infty$, so still $VSS^u \geq 0$.

Define finally *minimized expected objective function value of pairs subproblems* with fixed reference scenario ($MEPS^u$) as

$$MEPS^u = \min_{\{k=1,\ldots,K\} \cup \{u\}} \mathrm{E}_\xi z(\hat{x}^{u,k}, \xi),$$

where $\hat{x}^{u,k}$ is an optimal first stage solution of the pairs subproblem of ξ^u and ξ^k. Relationships between the new characteristics are derived in the following theorems.

Theorem 2.36 For the stochastic problem (2.29), for any reference scenario ξ^u, there is

$$RP \leq MEPS^u \leq EVRS^u.$$

Proof. RP, $MEPS^u$ and $EVRS^u$ are optimal values of objective function in the problem $\min_x \mathrm{E}_\xi z(x, \xi)$. When computing RP, we minimize over all $x \in F_1$ where F_1 is a set of all feasible solutions of (2.29). When computing $MEPS^u$, we consider as feasible all $x \in F_2 = F_1 \cap \{\hat{x}^{u,1}, \ldots, \hat{x}^{u,K}, \hat{x}^{u,u}\}$, and when computing $EVRS^u$, we have a feasibility set $F_3 = F_2 \cap \{\bar{x}^u\} = F_2 \cap \{\hat{x}^{u,u}\}$, since $\bar{x}^u = \hat{x}^{u,u}$.
The feasibility sets for RP, $MEPS^u$ and $EVRS^u$ are smaller and smaller, i.e. it holds $F_3 \subseteq F_2 \subseteq F_1$ and thus $RP \leq MEPS^u \leq EVRS^u$. □

We can obtain bounds on the value of stochastic solution from these new variables:

Theorem 2.37 For a stochastic program of the form (2.29), for any reference scenario ξ^u, there is

$$0 \leq EVRS^u - MEPS^u \leq VSS^u \leq EVRS^u - EOPS^u \leq EVRS^u - WS.$$

Proof. It follows immediately from theorems 2.35 and 2.36. $\square$
Remark 2.38 It is obvious that a sufficient condition for $EVPI = 0$ is the existence of x^* and $y^*(\xi^i)$ such that for every particular scenario ξ^i the pair $(x^*, y^*(\xi^i))$ is an optimal solution of the problem $\min\limits_x z(x, \xi^i)$ (i.e. first stage decision of the wait-and-see solution is identical for all scenarios).
To obtain a sufficient condition for $VSS^u = 0$, consider a pairs subproblem of ξ^u and ξ^k for $k \in \{1, \ldots, K\}$. If $(x^*, y^*(\xi^u))$ is an optimal solution of the problem

$$\begin{aligned} &\min_x z(x, \xi^u) = \min_x \left\{ c^\top x + \min_y \{ q^\top y : Wy = \xi^u - Tx, y \geq 0 \} \right\} \\ &\text{s.t. } Ax = b, \ x \geq 0 \end{aligned} \tag{2.32}$$

and for all $k = 1, \ldots, K$ there exists some $y^*(\xi^k)$ such that $(x^* = \hat{x}^{u,k}, y^*(\xi^u), y^*(\xi^k))$ solves the pairs subproblem of ξ^u and ξ^k, then x^* is the first stage optimal solution of the recourse problem (2.29). Then

$$\begin{aligned} EVRS^u &= \mathrm{E}_\xi z(\hat{x}^{u,k}, \xi) \ \forall k \\ &= \mathrm{E}_\xi z(x^*, \xi) \end{aligned}$$

and

$$RP = \min_x \mathrm{E}_\xi z(x, \xi) = \mathrm{E}_\xi z(x^*, \xi).$$

Hence, $EVRS^u = RP$ and $VSS^u = EVRS^u - RP = 0$.
We have already mentioned that $VSS = VSS^u$ for $\xi^u = \mathrm{E}\xi$, which can be used to find a sufficient condition for $VSS = 0$ as a special case of $VSS^u = 0$.

2.5. Group subproblem

We can generalize the idea of pairs subproblems, $MEPS^u$ and $EOPS^u$. Using larger groups of scenarios instead of pairs, we can obtain tighter bounds on the value of stochastic solution.
Consider the problem (2.29), where $\xi = h(\xi)$ is the only random variable, the support Ξ of ξ is finite, $\Xi = \{\xi^1, \ldots, \xi^K\}$ and for $k = 1, \ldots, K$ we denote $p^k = P(\xi = \xi^k)$. Consider a reference scenario ξ^u with $P(\xi^u) = p^u$. Define

$$EOGS^u(l) = \frac{1}{(1-p^u)^l} \sum_{\substack{i_1=1 \\ \xi^{i_1} \neq \xi^u}}^{K} \sum_{\substack{i_2 \geq i_1 \\ \xi^{i_2} \neq \xi^u}}^{K} \cdots \sum_{\substack{i_l \geq i_{l-1} \\ \xi^{i_l} \neq \xi^u}}^{K} p^{i_1} p^{i_2} \ldots p^{i_l} . \min z^l(x, \xi^u, \xi^{i_1}, \ldots, \xi^{i_l}),$$

where $z^l(x, \xi^u, \xi^{i_1}, \ldots, \xi^{i_l}) = c^\top x + p^u q^\top y(\xi^u) + \sum\limits_{i_h \in L_l} (1-p^u) q^\top y(\xi^{i_h})$ is an objective function of the group(l) subproblem

$$\begin{aligned} \min_{(x, y(\xi^u), y(\xi^{i_1}), \ldots, y(\xi^{i_l}))} \quad & z^l(x, \xi^u, \xi^{i_1}, \ldots, \xi^{i_l}) \\ \text{s.t.} \quad & Ax = b, \ x \geq 0, \\ & Wy(\xi^u) = \xi^u - Tx, \\ & Wy(\xi^{i_h}) = \xi^{i_h} - Tx, \ h = 1, \ldots, l, \\ & y(\xi^u), y(\xi^{i_h}) \geq 0, \ h = 1, \ldots, l, \end{aligned} \tag{2.33}$$

where L_l is the set of distinct indices among $i_1, \dots, i_l$.
In the definition of $EOGS^u(l)$ we sum up over all bags[3] with l elements chosen from the set $\{\xi^1, \dots, \xi^K\} \setminus \{\xi^u\}$. It could seem that we sum up over all nondecreasing sequences with elements in $\{\xi^1, \dots, \xi^K\} \setminus \{\xi^u\}$, but it does not depend on ordering of the elements. So we can't assume that all the indices among $i_1, \dots, i_l$ are distinct since we compute the min $z^l(x, \xi^{i_1}, \dots, \xi^{i_l})$ for all bags, not sets, of scenarios.
Define *minimized expectation of the objective function value with fixed optimal solution of group(l) subproblems* ($MEGS^u(l)$) with fixed reference scenario ξ^u: If $\hat{x}(\xi^u, \xi^{i_1}, \dots, \xi^{i_l})$ is an optimal solution of the group(l) subproblem (2.33), then

$$MEGS^u(l) = \min_{(i_1,\dots,i_l)} \mathrm{E}_\xi[z(\hat{x}(\xi^u, \xi^{i_1}, \dots, \xi^{i_l}), \xi)] = \min_{(i_1,\dots,i_l)} \sum_{k=1}^{K} p^k z(\hat{x}(\xi^u, \xi^{i_1}, \dots, \xi^{i_l}), \xi^k).$$

According to the definition of $MEGS^u(l)$ we consider every group(l) subproblem with one fixed reference scenario ξ^u, find its optimal solution $\hat{x}(\xi^u, \xi^{i_1}, \dots, \xi^{i_l})$ and compute the true expectation (over all scenarios) of the objective function value $z(\hat{x}(\xi^u, \xi^{i_1}, \dots, \xi^{i_l}), \xi)$. Finally, we minimize this over all bags with l elements.
Theorem 2.39 For the two stage stochastic program of the form (2.29) and for any fixed ξ^u it holds

$$WS \leq EOPS^u = EOGS^u(1) \leq EOGS^u(2) \leq \ldots \leq EOGS^u(K-2) \leq EOGS^u(K-1) \leq RP.$$

Proof. $EOGS^u(1) = EOPS^u$ according to their definitions.
Let $(\bar{x}, \bar{y}(\xi^u), \bar{y}(\xi^{i_1}), \dots, \bar{y}(\xi^{i_q}))$ be an optimal solution of the group(q) subproblem with the reference scenario ξ^u.
Let $(x^{*|i_{q+1}=j}, y^{*|i_{q+1}=j}(\xi^u), y^{*|i_{q+1}=j}(\xi^{i_1}), \dots, y^{*|i_{q+1}=j}(\xi^{i_q}))$ be an optimal solution (without the last component) of group($q+1$) subproblem with the reference scenario ξ^u under condition that $i_{q+1} = j$ for $j \in \{1, \dots, K\}$.
Now we will define a vector of vectors:

$$\begin{aligned} &(\hat{x}, \hat{y}(\xi^u), \hat{y}(\xi^{i_1}), \dots, \hat{y}(\xi^{i_q})) = \\ &= \frac{1}{1-p^u} \sum_{\substack{j=1 \\ \xi^j \neq \xi^u}}^{K} p^j . \left(x^{*|i_{q+1}=j}, y^{*|i_{q+1}=j}(\xi^u), y^{*|i_{q+1}=j}(\xi^{i_1}), \dots, y^{*|i_{q+1}=j}(\xi^{i_q})\right) \end{aligned} \quad (2.34)$$

in the sense that

$$\hat{x}_i = \frac{1}{1-p^u} \sum_{\substack{j=1 \\ \xi^{i_j} \neq \xi^u}}^{K} p^j . x_i^{*|i_{q+1}=j}$$

for all components $\hat{x}_i$ of $\hat{x}$, and similarly for the other vectors $\hat{y}(\xi^u), \hat{y}(\xi^{i_1}), \dots, \hat{y}(\xi^{i_q})$.
Since $(\bar{x}, \bar{y}(\xi^u), \bar{y}(\xi^{i_1}), \dots, \bar{y}(\xi^{i_q}))$ is an optimal solution of the group(q) subproblem and $(\hat{x}, \hat{y}(\xi^u), \hat{y}(\xi^{i_1}), \dots, \hat{y}(\xi^{i_q}))$ is feasible for the same problem, we have

---

[3] Bag is a system of elements in which we do not care for the ordering of elements (similarly in a set) and elements in a bag need not be distinct (unlike in a set). E.g. [1, 1, 2, 2] is the same bag as [1, 2, 1, 2] but we cannot write it as [1, 1, 2] or [1, 2, 2] or [1, 2].

$$\min z^l(x,\xi^u,\xi^{i_1},\dots,\xi^{i_q}) =$$
$$= c^\top \bar{x} + p^u q^\top \bar{y}(\xi^u) + \sum_{i_h \in L_q} (1-p^u) q^\top \bar{y}(\xi^{i_h}) \leq c^\top \hat{x} + p^u q^\top \hat{y}(\xi^u) + \sum_{i_h \in L_q} (1-p^u) q^\top \hat{y}(\xi^{i_h}).$$

Adding probabilities and nested sums, we obtain

$$\sum_{\substack{i_1=1\\ \xi^{i_1}\neq\xi^u}}^{K} \cdots \sum_{\substack{i_q\geq i_{q-1}\\ \xi^{i_q}\neq\xi^u}}^{K} p^{i_1}\dots p^{i_q}\left[c^\top \bar{x} + p^u q^\top \bar{y}(\xi^u) + \sum_{i_h \in L_q} (1-p^u) q^\top \bar{y}(\xi^{i_h})\right] \leq$$
$$\leq \sum_{\substack{i_1=1\\ \xi^{i_1}\neq\xi^u}}^{K} \cdots \sum_{\substack{i_q\geq i_{q-1}\\ \xi^{i_q}\neq\xi^u}}^{K} p^{i_1}\dots p^{i_q}\left[c^\top \hat{x} + p^u q^\top \hat{y}(\xi^u) + \sum_{i_h \in L_q} (1-p^u) q^\top \hat{y}(\xi^{i_h})\right].$$

Dividing by $(1-p^u)^q$ and then rewriting vectors according to equation (2.34) we obtain

$$EOGS^u(q) =$$
$$= \frac{1}{(1-p^u)^q} \sum_{\substack{i_1=1\\ \xi^{i_1}\neq\xi^u}}^{K} \cdots \sum_{\substack{i_q\geq i_{q-1}\\ \xi^{i_q}\neq\xi^u}}^{K} p^{i_1}\dots p^{i_q}\left[c^\top \bar{x} + p^u q^\top \bar{y}(\xi^u) + \sum_{i_h \in L_q} (1-p^u) q^\top \bar{y}(\xi^{i_h})\right]$$
$$\leq \frac{1}{(1-p^u)^q} \sum_{\substack{i_1=1\\ \xi^{i_1}\neq\xi^u}}^{K} \cdots \sum_{\substack{i_q\geq i_{q-1}\\ \xi^{i_q}\neq\xi^u}}^{K} p^{i_1}\dots p^{i_q}\left[c^\top \hat{x} + p^u q^\top \hat{y}(\xi^u) + \sum_{i_h \in L_q} (1-p^u) q^\top \hat{y}(\xi^{i_h})\right]$$
$$= \frac{1}{(1-p^u)^q} \sum_{\substack{i_1=1\\ \xi^{i_1}\neq\xi^u}}^{K} \cdots \sum_{\substack{i_q\geq i_{q-1}\\ \xi^{i_q}\neq\xi^u}}^{K} p^{i_1}\dots p^{i_q}\left[c^\top \left(\frac{1}{1-p^u} \sum_{\substack{j=1\\ \xi^j\neq\xi^u}}^{K} p^j x^{*|i_{q+1}=j}\right) + \right.$$
$$\left. + p^u q^\top \left(\frac{1}{1-p^u} \sum_{\substack{j=1\\ \xi^j\neq\xi^u}}^{K} p^j y^{*|i_{q+1}=j}(\xi^u)\right) + \sum_{i_h \in L_q} (1-p^u) q^\top \left(\frac{1}{1-p^u} \sum_{\substack{j=1\\ \xi^j\neq\xi^u}}^{K} y^{*|i_{q+1}=j}(\xi^{i_h})\right)\right]$$
$$= \frac{1}{(1-p^u)^{q+1}} \sum_{\substack{i_1=1\\ \xi^{i_1}\neq\xi^u}}^{K} \cdots \sum_{\substack{i_q\geq i_{q-1}\\ \xi_{i_q}\neq\xi^u}}^{K} \sum_{\substack{j=1\\ \xi^j\neq\xi^u}}^{K} p^{i_1}\dots p^{i_q} p^j \Big[c^\top x^{*|i_{q+1}=j} + p^u q^\top y^{*|i_{q+1}=j}(\xi^u) +$$
$$+ \sum_{i_h \in L_{q+1}} (1-p^u) q^\top y^{*|i_{q+1}=j}(\xi^{i_h})\Big]$$
$$= EOGS^u(q+1).$$

Comments: $x^{*|i_{q+1}=j}, y^{*|i_{q+1}=j}(\xi^u), y^{*|i_{q+1}=j}(\xi^1), \dots, y^{*|i_{q+1}=j}(\xi^{i_j})$ are relevant components of the optimal solution of the group$(q+1)$ subproblem with $\xi^{i_q+1} = \xi^j$. We have summed up over this last scenario as well, so the last $(q+1)$-fold sum includes all sequences of the length $(q+1)$ created with numbers from the set $\{1,\dots,K\}\setminus\{u\}$. First q components of this sequences create nondecreasing sequences and the $(q+1)$-th component is added after it. But in fact we do not care about ordering of the elements in these sequences, because according to the definition of $EOGS^u(q+1)$ we have to sum up over all bags of $(q+1)$ numbers chosen from the set $\{1,\dots,K\}\setminus\{u\}$, and we really do so. So the last expression is really $EOGS^u(q+1)$ according to its definition.

Also, components of any feasible solution of the whole recourse problem will form a feasible solution of group$(K-1)$ subproblem. Hence, $EOGS^u(K-1)\le RP$. □

Theorem 2.40 For the stochastic program of the form (2.29) and for any fixed reference scenario ξ^u it holds

$$RP\le MEGS^u(K-1)\le MEGS^u(K-2)\le\ldots\le MEGS^u(2)\le MEGS^u(1)=MEPS^u\le EEV.$$

Proof. $MEGS^u(1)=MEPS^u$ by definitions. For any $l=2,\dots,K$, any solution considered in $MEGS^u(l-1)$ is also considered in $MEGS^u(l)$. Hence, $MEGS^u(l)$ cannot be worse (i.e. greater) than $MEGS^u(l-1)$, i.e. $MEGS^u(l)\le MEGS^u(l-1)$. Also, any first stage solution of group$(K-1)$ subproblem can be completed to the feasible solution of the whole recourse problem and so we obtain $MEGS^u(K-1)\ge RP$.
The inequality $MEPS^u\le EEV$ holds according to theorem 2.36 thanks to the fact that $EVRS^u=EEV$ for reference scenario $\xi^u=\mathrm{E}\xi$. □

Theorem 2.41 For a stochastic program of the form (2.29), for any fixed ξ^u, there is

$$\begin{gathered}0\le EEV-MEPS^u=EEV-MEGS^u(1)\le EEV-MEGS^u(2)\le\ldots\le\\ \le EEV-MEGS^u(K-1)\le VSS\le EEV-EOGS^u(K-1)\le EEV-EOGS^u(K-2)\le\\ \ldots\le EEV-EOGS^u(2)\le EEV-EOGS^u(1)=EEV-EOPS^u\le EEV-WS.\end{gathered}$$

Proof. It follows immediately from theorems 2.39 and 2.40. □

However, computing $MEGS^u(l)$ and $EOGS^u(l)$ for larger groups may be more difficult than computing the here-and-now solution. Solving the recourse problem is of equivalent size as solving a group$(K-1)$ subproblem. Therefore, the bounds shown in theorem 2.41 are worth only if they bring information to the problem without being too computationally demanding. For this reason, the group subproblems have not become too popular.

2.6. Modified wait-and-see approach

This approach, slightly different from the wait-and-see approach, has been established in [7]. It emphasises on computational efficiency and usefulness in particular real-life applications. We will follow the approach presented in [7] and then we will generalize it. We will also show that the original approach can be seen as incorrect from a little different point of view.

Let us have a standard two-stage decision problem

$$\begin{aligned}&\min_x \mathrm{E}_\xi z(x,\xi)=\min_x\left\{c^Tx+\mathrm{E}_\xi\min_y\{q^T(\xi)y: W(\xi)y=h(\xi)-T(\xi)x, y\ge 0\}\right\}\\ &\text{s.t. } Ax=b,\ x\ge 0\end{aligned}\tag{2.35}$$

and suppose that the support Ξ of ξ is finite, scenarios $\xi^1,\dots,\xi^K$ can realize. Denote again $p^k=P(\xi=\xi^k)$ for $k=1,\dots,K$.

According to [7], from an application viewpoint, the here-and-now solution is seen as maybe a little doubtful. Any first stage wait-and-see solution $x(\xi^s)$ obtained as a solution of (2.10) is optimal for at least one scenario (the ξ^s), but no such optimality is guaranteed for the here-and-now solution x^*. So an analysis of alternatives and their consequent objective function values is suggested to be done with help of so called modified wait-and-see solutions.

For $i = 1, \ldots, K$ we denote as $x^*_{\{i\}}$ an optimal first stage solution of the problem for the scenario ξ^i

$$\begin{aligned} &\min_x \; z(x, \xi^i) = \min_x \left\{ c^T x + \min_y \{ q^T(\xi^i) y : W(\xi^i) y = h(\xi^i) - T(\xi^i) x, y \geq 0 \} \right\} \\ &\text{s.t. } Ax = b, \; x \geq 0, \end{aligned} \tag{2.36}$$

and suppose (or consider only such i) that $-\infty < z(x^*_{\{i\}}, \xi^k) < +\infty$ for $k = 1, \ldots, K$. This means that the second stage solution found as minimizing with the first stage decision fixed as $x^*_{\{i\}}$ and after realization ξ^k is feasible and leads to a bounded objective function value.

We define the *modified wait-and-see value* related to the i-th scenario as

$$\begin{aligned} MWS_i &= \mathrm{E}_\xi z(x^*_{\{i\}}, \xi) \\ &= c^T x^*_{\{i\}} + \sum_{k=1}^{K} p^k . \min_y \left\{ q(\xi^k)^T y : W(\xi^k) y = h(\xi^k) - T(\xi^k) x^*_{\{i\}}, y \geq 0 \right\}. \end{aligned} \tag{2.37}$$

The wait-and-see objective function value for realization ξ^i is $z(x^*_{\{i\}}, \xi^i)$ and it is clear that $z(x^*_{\{i\}}, \xi^i) \leq z(x^*_{\{k\}}, \xi^i)$ for all $k = 1, \ldots, K$.

We have defined the optimal objective function value of the here-and-now problem as $RP = \min_x \mathrm{E}_\xi z(x, \xi)$ and so $RP \leq MWS_i$ for all $i = 1, \ldots, K$. Hence, there is

$$U := \min_{\{i : \xi^i \in \Xi\}} MWS_i \geq RP \geq WS =: L \tag{2.38}$$

and we can define *expected value of perfect information for the modified wait-and-see approach* as

$$MEVPI = U - L = \min_{\{i : \xi^i \in \Xi\}} MWS_i - WS.$$

It is clear that $MEVPI \geq EVPI$ according to the definition of U and L and so we have a new (easily evaluable) upper bound on $EVPI$. An upper bound for the cost of uncertainty for each $x^*_{\{i\}}$ is $MWS_i - WS$, which is nonnegative.

According to [7], modified wait-and-see analysis suggested for applications is as follows:

1. Find $x^*_{\{i\}}$ for (not necessarily) all $i = 1, \ldots, K$.
2. Compute $z(x^*_{\{i\}}, \xi^k), k = 1, \ldots, K$ for all distinct $x^*_{\{i\}}$s.
3. Compare these solutions to decide, which $x^*_{\{i\}}$s are good, acceptable, risky or totally bad. This step is discussed more in the next paragraph.

Let us discuss advantages of the modified wait-and-see (MWS) approach as compared to the here-and-now approach.

The first advantage of the modified wait-and-see approach is an easy detection of the source of infeasibility. When infeasibility appears in a here-and-now problem, it is not easy to find the scenarios which cause it. This situation can arise when solving a problem with incomplete recourse. When applying the MWS analysis it is clear that if there is no feasible solution for a scenario ξ^i, then any here-and-now model which includes ξ^i would be problematic. It is also easy to find which constraints are violated. The "state of the world" causing the problem is obvious, since only one state of the world is involved in the given MWS problem.

In the suggested MWS "what-if analysis", we compute $z(x^*_{\{i\}}, \xi^k)$ for all pairs (ξ^i, ξ^k) of scenarios. Hence, we can see which results can be expected for which realizations ξ^k when using $x^*_{\{i\}}$ as the first stage decision. This decision is naturally optimal in the case that ξ^i really happens, but it may be much worse when some of the other scenarios realize. There can exist a scenario $\xi^j \in \Xi$ such that $z(x^*_{\{i\}}, \xi^j)$ is extremely large. Intuitively, if the probability $p^j = P(\xi = \xi^j)$ is small and the decision $x^*_{\{i\}}$ gives "good results" for the other scenarios (except for ξ^j), it can happen that the optimal first stage here-and-now solution is $x^* = x^*_{\{i\}}$. This minimizes the expected objective function value, but it also leads to the extremely large costs $z(x^*_{\{i\}}, \xi^j)$ with probability p^j. Sometimes we do not want to undergo such risk and we find better to choose another first stage decision, which is not optimal for the here-and-now problem, but which is safer for all scenarios (especially for ξ^j).

Via MWS approach, we can decide (before solving the here-and-now problem) which solutions among $x^*_{\{i\}},\ i = 1, \ldots, K$ are good and which of them should be rejected even though they were optimal for the here-and-now problem. Moreover, the MWS approach generates complete probability distribution of the objective function values (costs) for the various decision alternatives $x^*_{\{i\}},\ i = 1, \ldots, K$, so we can rank the alternatives by expected objective function values and we can see which first stage decisions would be good for all scenarios, which would be good for most of them and risky for the others, and which would be bad with unacceptably large probability.

For comparison, in a common scenario analysis, we compute $z(x^*, \xi^k),\ k = 1, \ldots, K$, for x^* an optimal first-stage RP solution, which also leads to revealing the scenarios that would bring too large costs. However, the scenario analysis does not offer any other alternative if we decide to reject x^*.

Note also that computational demand for MWS_i for all i is about K-times less than that required to obtain the here-and-now solution.

2.6.1. Generalized modified wait-and-see approach

A straightforward generalization of the MWS approach leads to computing the optimal first-stage solution of the problem for some subset of possible scenarios, while keeping proportions of their a priori probabilities. Using larger and larger subsets we gain better approximation of the true distribution of the random variable ξ, so a chain of inequalities between optimal objective function values (of the problems with particular subsets of scenarios) could be expected. However, we will show that in general no such inequalities work.

Consider the original problem (2.35) again. The support Ξ of ξ is finite, $\Xi = \{\xi^1, \ldots, \xi^K\}$ and $P(\xi^k) = p^k$ for $k = 1, \ldots, K$.

Now, we obtain a (partial) information that one of two scenarios ξ^i, ξ^j will happen (but we do not know which one of them). We then solve the problem for the two scenarios only:

$$\begin{aligned}\min_x \Big\{ & c^T x + \frac{p^i}{p^i + p^j} . \min_y \{ q(\xi^i)^T y : W(\xi^i) y = h(\xi^i) - T(\xi^i) x,\ y \geq 0 \} \\ & + \frac{p^j}{p^i + p^j} . \min_y \{ q(\xi^j)^T y : W(\xi^j) y = h(\xi^j) - T(\xi^j) x,\ y \geq 0 \} \Big\} \\ \text{s.t. } & Ax = b,\ x \geq 0.\end{aligned}$$

Denote $x^*_{\{i,j\}}$ an optimal first stage decision of this problem. Suppose for simplicity that for all pairs (i, j) the value $z(x^*_{\{i,j\}}, \xi^k)$ is finite for all $k = 1, \ldots, K$.
Let us define

$$\begin{aligned} MWS_{\{i,j\}} &= \mathrm{E}_\xi z(x^*_{\{i,j\}}, \xi) \\ &= c^T x^*_{\{i,j\}} + \sum_{k=1}^{K} p^k . \min_y \left\{ q(\xi^k)^T y : W(\xi^k) y = h(\xi^k) - T(\xi^k) x^*_{\{i,j\}},\ y \geq 0 \right\}, \end{aligned}$$

i.e. $MWS_{\{i,j\}}$ denotes expected costs when $x^*_{\{i,j\}}$ is used as the first stage solution.
Define

$$MEVPI_2 = \min_{\substack{\{i,j\} \subseteq \{1,\ldots,K\} \\ i \neq j}} MWS_{\{i,j\}} - WS.$$

The $x^*_{\{i,j\}}$ is feasible for the original problem (2.35) which implies that $MWS_{\{i,j\}} \geq RP$ for all (i, j) and so

$$U_2 := \min_{\substack{\{i,j\} \subseteq \{1,\ldots,K\} \\ i \neq j}} MWS_{\{i,j\}} \geq RP \geq WS =: L$$

which implies that

$$U_2 - L = \min_{\substack{\{i,j\} \subseteq \{1,\ldots,K\} \\ i \neq j}} MWS_{\{i,j\}} - WS = MEVPI_2 \geq RP - WS = EVPI.$$

Now we can introduce the generalization for N-tuples, $N < K$. In the following, we will deal with finite sets S only and we will denote as $\#S$ the number of elements of the set S. The problem for scenarios with indices in the selected subset $S \subseteq \{1, \ldots, K\}$, $\#S = N$, reads:

$$\begin{aligned}\min_x & \left\{ c^T x + \sum_{n \in S} \frac{p^n}{\sum_{l \in S} p^l} \min_y \left\{ q(\xi^n)^T y : W(\xi^n) y = h(\xi^n) - T(\xi^n) x,\ y \geq 0 \right\} \right\} \\ \text{s.t. } & Ax = b,\ x \geq 0,\end{aligned}$$

and the optimal first stage solution is denoted as x^*_S.
We fix this x^*_S as the first stage solution and we compute the actual expectation of the optimal objective function value (i.e. expected costs):

$$MWS_S = \mathrm{E}_\xi z(x^*_S, \xi) = c^T x^*_S + \sum_{k=1}^{K} p^k \min_y \left\{ q(\xi^k)^T y : W(\xi^k) y = h(\xi^k) - T(\xi^k) x^*_S,\ y \geq 0 \right\}.$$

The x_S^* is feasible for the original problem (2.35), and so $MWS_S \geq RP$ for every set $S \subseteq \{1,\dots,K\}$ such that $\#S = N$. Hence,

$$U_N := \min_{\substack{S\subseteq\{1,\dots,K\}\\ \#S=N}} MWS_S \geq RP \geq WS =: L$$

which implies that

$$U_N - L = \min_{\substack{S\subseteq\{1,\dots,K\}\\ \#S=N}} MWS_S - WS = MEVPI_N \geq RP - WS = EVPI,$$

where we defined

$$MEVPI_N = \min_{\substack{S\subseteq\{1,\dots,K\}\\ \#S=N}} MWS_S - WS.$$

For two sets $S \subseteq \{1,\dots,K\}$ and $Z \subseteq \{1,\dots,K\}$ such that $\#S = N$ and $\#Z = N+1$, we would like to know whether any inequality works between MWS_S and MWS_Z, or $\min\limits_S MWS_S$ and $\min\limits_Z MWS_Z$. It is quite easy to see that no such inequality can hold true in general. There are two (quite instructive) counterexamples.

Example 2.42 At first, the problem is solved for scenarios $\xi^n,\ n \in S$. So the decision maker hedges against $N = \#S$ scenarios only. When solving the problem for the subset $Z,\ \#Z = N+1$, he hedges against $N+1$ scenarios, which seems to be better. So we can expect $MWS_Z \leq MWS_S$, which, however, need not be true.
As an example, suppose $K = 3$ (number of scenarios), $\#S = 1$, $\#Z = 2$, probability of each scenario is $\frac{1}{3}$, the only randomness is in the second stage costs:

$$q(\xi^1) = \begin{pmatrix}1000\\1\\1\end{pmatrix}, q(\xi^2) = \begin{pmatrix}1\\1000\\1\end{pmatrix}, q(\xi^3) = \begin{pmatrix}1\\1\\1000\end{pmatrix}; c = \begin{pmatrix}10\\10\\10\end{pmatrix}.$$

The form of the recourse problem is as follows:

$$\begin{aligned}&\min_x \left[\begin{pmatrix}10\\10\\10\end{pmatrix}^T x + \frac{1}{3}\min_y\left\{\begin{pmatrix}1000\\1\\1\end{pmatrix}^T y : y_i \geq 1 - x_i,\ i = 1,2,3\right\} + \right.\\ &\left. +\frac{1}{3}\min_y\left\{\begin{pmatrix}1\\1000\\1\end{pmatrix}^T y :\ y_i \geq 1 - x_i,\ i = 1,2,3\right\} + \frac{1}{3}\min_y\left\{\begin{pmatrix}1\\1\\1000\end{pmatrix}^T y : y_i \geq 1 - x_i,\ i = 1,2,3\right\}\right]\\ &\text{ s.t. } 0 \leq x_i \leq 1,\ i = 1,2,3.\end{aligned}$$

Now we will compute $MWS_{\{1\}}$:

$$\begin{aligned}&\min_x \left[\begin{pmatrix}10\\10\\10\end{pmatrix}^T x + \min_y\left\{\begin{pmatrix}1000\\1\\1\end{pmatrix}^T y : y_i \geq 1 - x_i,\ i = 1,2,3\right\}\right]\\ &\text{ s.t. } 0 \leq x_i \leq 1,\ i = 1,2,3,\end{aligned}$$

which gives an optimal first stage solution $x_{\{1\}}^* = (1,0,0)^T$.

$$\begin{aligned}MWS_{\{1\}} &= \begin{pmatrix}10\\10\\10\end{pmatrix}^T \begin{pmatrix}1\\0\\0\end{pmatrix} + \frac{1}{3}\min_y\left\{\begin{pmatrix}1000\\1\\1\end{pmatrix}^T y : y_i \geq 1 - x_{\{1\}i}^*,\ i = 1,2,3\right\} + \\ &+\frac{1}{3}\min_y\left\{\begin{pmatrix}1\\1000\\1\end{pmatrix}^T y : y_i \geq 1 - x_{\{1\}i}^*,\ i = 1,2,3\right\} + \frac{1}{3}\min_y\left\{\begin{pmatrix}1\\1\\1000\end{pmatrix}^T y : y_i \geq 1 - x_{\{1\}i}^*,\ i = 1,2,3\right\}\\ &= 10 + \frac{1}{3}(1+1) + \frac{1}{3}(1000+1) + \frac{1}{3}(1+1000) = 678.\end{aligned}$$

Thanks to the total symmetry of the problem, we have

$$MWS_{\{1\}} = MWS_{\{2\}} = MWS_{\{3\}} = 678.$$

Now we will compute $MWS_{\{1,2\}}$:

$$\min_x \left[\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}^T x + \frac{1}{2} \min_y \left\{ \begin{pmatrix} 1000 \\ 1 \\ 1 \end{pmatrix}^T y : y_i \geq 1 - x_i,\ i = 1,2,3 \right\} + \right.$$
$$\left. + \frac{1}{2} \min_y \left\{ \begin{pmatrix} 1 \\ 1000 \\ 1 \end{pmatrix}^T y : y_i \geq 1 - x_i,\ i = 1,2,3 \right\} \right]$$
$$\text{s.t. } 0 \leq x_i \leq 1,\ i = 1,2,3$$

and the optimal first stage solution is $x^*_{\{1,2\}} = (1,1,0)^T$.

$$MWS_{\{1,2\}} = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}^T \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1000 \\ 1 \\ 1 \end{pmatrix}^T y : y_i \geq 1 - x^*_{\{1,2\}i},\ i = 1,2,3 \right\} +$$
$$+ \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1 \\ 1000 \\ 1 \end{pmatrix}^T y : y_i \geq 1 - x^*_{\{1,2\}i},\ i = 1,2,3 \right\} + \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1 \\ 1 \\ 1000 \end{pmatrix}^T y : y_i \geq 1 - x^*_{\{1,2\}i},\ i = 1,2,3 \right\}$$
$$= 10 + 10 + \frac{1}{3} + \frac{1}{3} + \frac{1}{3}.1000 = 354.$$

Again, because of the symmetry of the problem, there is

$$MWS_{\{1,2\}} = MWS_{\{1,3\}} = MWS_{\{2,3\}} = 354.$$

In this example, we obtained the expected result

$$\min_{\{i\} \subseteq \{1,2,3\}} MWS_{\{i\}} > \min_{\substack{\{i,j\} \subseteq \{1,2,3\} \\ i \neq j}} MWS_{\{i,j\}}.$$

Example 2.43 The result changes when the price of hedging in the first stage is so high that the second stage recourse cannot compensate it and it is better not to hedge against some scenarios. This situation results in $MWS_S \leq MWS_Z$.
As an example, suppose $K = 3$ (number of all possible scenarios), $\#S = 1$ and $\#Z = 2$, $q = q(\xi)$ is the only random element and there is

$$q(\xi^1) = \begin{pmatrix} -9000 \\ 1 \\ 1 \end{pmatrix},\ q(\xi^2) = \begin{pmatrix} 1 \\ -4004 \\ 1 \end{pmatrix},\ q(\xi^2) = \begin{pmatrix} 1 \\ 1 \\ -4004 \end{pmatrix};\ c = \begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix},$$

and the constraints are of the form $0 \leq x_i \leq 2M_i$ for $i = 1,2,3$ (M_i large positive, e.g. $M_i = 1000000$ for $i = 1,2,3$) in the first stage, and $0 \leq y_i \leq \frac{1}{2}x$ for $i = 1,2,3$ in the second stage. Probability of each scenario is $\frac{1}{3}$. The recourse problem is then

$$\min_x \left[\begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix}^T x + \frac{1}{3} \min_{0 \leq y \leq \frac{1}{2}x} \left\{ \begin{pmatrix} -9000 \\ 1 \\ 1 \end{pmatrix}^T y \right\} + \right.$$
$$\left. + \frac{1}{3} \min_{0 \leq y \leq \frac{1}{2}x} \left\{ \begin{pmatrix} 1 \\ -4004 \\ 1 \end{pmatrix}^T y \right\} + \frac{1}{3} \min_{0 \leq y \leq \frac{1}{2}x} \left\{ \begin{pmatrix} 1 \\ 1 \\ -4004 \end{pmatrix}^T y \right\} \right]$$
$$\text{s.t. } 0 \leq x \leq 2M$$

for $M = (M_1, M_2, M_3)^\top$. For simplicity of writing, we shall not write the constraints, they are kept in the same fixed shape as above.
Now we will compute $MWS_{\{1\}}$. The first problem reads

$$\min_x \left[\begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix}^\top x + \min_y \begin{pmatrix} -9000 \\ 1 \\ 1 \end{pmatrix}^\top y \right]$$

which results in the first stage optimal solution $x^*_{\{1\}} = (2M_1, 0, 0)^\top$. Then

$$\begin{aligned} MWS_{\{1\}} &= \begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix}^\top \begin{pmatrix} 2M_1 \\ 0 \\ 0 \end{pmatrix} + \left[\frac{1}{3} \min_y \begin{pmatrix} -9000 \\ 1 \\ 1 \end{pmatrix}^\top y + \frac{1}{3} \min_y \begin{pmatrix} 1 \\ -4004 \\ 1 \end{pmatrix}^\top y + \frac{1}{3} \min_y \begin{pmatrix} 1 \\ 1 \\ -4004 \end{pmatrix}^\top y \right] \\ &= 2000M_1 + \frac{1}{3}(-9000M_1) = -1000M_1. \end{aligned}$$

Now we will compute $MWS_{\{1,2\}}$:

$$\min_x \begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix}^\top x + \frac{1}{2} \min_y \begin{pmatrix} -9000 \\ 1 \\ 1 \end{pmatrix}^\top y + \frac{1}{2} \min_y \begin{pmatrix} 1 \\ -4004 \\ 1 \end{pmatrix}^\top y. \tag{2.39}$$

The first stage optimal solution is $x^*_{\{1,2\}} = (2M_1, 2M_2, 0)^\top$.

$$\begin{aligned} MWS_{\{1,2\}} &= \begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix}^\top \begin{pmatrix} 2M_1 \\ 2M_2 \\ 0 \end{pmatrix} + \left[\frac{1}{3} \min_y \begin{pmatrix} -9000 \\ 1 \\ 1 \end{pmatrix}^\top y + \frac{1}{3} \min_y \begin{pmatrix} 1 \\ -4004 \\ 1 \end{pmatrix}^\top y + \frac{1}{3} \min_y \begin{pmatrix} 1 \\ 1 \\ -4004 \end{pmatrix}^\top y \right] \\ &= 2000M_1 + 2000M_2 + \frac{1}{3}(-9000M_1 - 4004M_2) = -1000M_1 + 665\frac{1}{3}M_2. \end{aligned} \tag{2.40}$$

Now we will compute $MWS_{\{1,3\}}$. At first, we solve

$$\min_x \left[\begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix}^\top x + \frac{1}{2} \min_y \begin{pmatrix} -9000 \\ 1 \\ 1 \end{pmatrix}^\top y + \frac{1}{2} \min_y \begin{pmatrix} 1 \\ 1 \\ -4004 \end{pmatrix}^\top y \right],$$

which results in the first stage optimal solution $x^*_{\{1,3\}} = (2M_1, 0, 2M_3)^\top$. Then

$$\begin{aligned} MWS_{\{1,3\}} &= \begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix}^\top \begin{pmatrix} 2M_1 \\ 0 \\ 2M_3 \end{pmatrix} + \left[\frac{1}{3} \min_y \begin{pmatrix} -9000 \\ 1 \\ 1 \end{pmatrix}^\top y + \frac{1}{3} \min_y \begin{pmatrix} 1 \\ -4004 \\ 1 \end{pmatrix}^\top y + \frac{1}{3} \min_y \begin{pmatrix} 1 \\ 1 \\ -4004 \end{pmatrix}^\top y \right] \\ &= 2000M_1 + 2000M_3 + \frac{1}{3}(-9000M_1 - 4004M_3) = -1000M_1 + 665\frac{1}{3}M_3. \end{aligned}$$

In the end, we will compute $MWS_{\{2,3\}}$:

$$\min_x \begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix}^\top x + \frac{1}{2} \min_y \begin{pmatrix} 1 \\ -4004 \\ 1 \end{pmatrix}^\top y + \frac{1}{2} \min_y \begin{pmatrix} 1 \\ 1 \\ -4004 \end{pmatrix}^\top y$$

gives optimal first stage solution $x^*_{\{2,3\}} = (0, 2M_2, 2M_3)^T$. Finally,

$$MWS_{\{2,3\}} = \begin{pmatrix}1000\\1000\\1000\end{pmatrix}^T \begin{pmatrix}0\\2M_2\\2M_3\end{pmatrix} + \left[\frac{1}{3}\min_y \begin{pmatrix}-9000\\1\\1\end{pmatrix}^T y + \frac{1}{3}\min_y \begin{pmatrix}1\\-4004\\1\end{pmatrix}^T y + \frac{1}{3}\min_y \begin{pmatrix}1\\1\\-4004\end{pmatrix}^T y\right]$$
$$= 2000M_2 + 2000M_3 + \frac{1}{3}(-4004M_2 - 4004M_3) = 665\frac{1}{3}M_2 + 665\frac{1}{3}M_3.$$

To summarize our results, we obtained

$$MWS_{\{1\}} = -1000M_1, \qquad MWS_{\{1,2\}} = -1000M_1 + 665\frac{1}{3}M_2,$$
$$MWS_{\{1,3\}} = -1000M_1 + 665\frac{1}{3}M_3, \qquad MWS_{\{2,3\}} = 665\frac{1}{3}M_2 + 665\frac{1}{3}M_3.$$

Therefore, we have

$$\min_{\{i\}\subseteq\{1,2,3\}} MWS_{\{i\}} \leq MWS_{\{1\}} < \min_{\substack{\{i,j\}\subseteq\{1,2,3\}\\ i\neq j}} MWS_{\{i,j\}}.$$

Let us analyse in short why the result is just this. When solving the first problem for any couple of i, j, we have the probabilities of each scenario equal to $\frac{1}{2}$, but in computing the second problem for the same couple of i, j, the recourse is added with weights of $\frac{1}{3}$. So, when solving the first problem (e.g. problem (2.39)), it is worth doing to "prepare ourselves" for the second stage (e.g., $\frac{1}{2}.(-4004)$ is less then the costs needed in the first stage), but when computing the second problem (e.g. problem (2.40)) with all three scenarios, it is not worth doing any more (e.g., $\frac{1}{3}.(-4004)$ is larger then the costs needed in the first stage). In fact, when solving the problem for two scenarios for the first time, the costs in the first stage are computed with different weights then in the second stage. This distortion is caused by norming.

2.6.2. Redefinition of MWS and generalized MWS approach

It seems logical to redefine the notion of the MWS_S in the sense which follows from the discussion of the previous example. Here, when solving the problem for a subset of scenarios, the costs in the first stage are computed with different weights then the costs in the second stage. We can change the norming in the following way.
Firstly, we will consider a case when $\#S = 1$. We solve the problem for one scenario ξ^i at first:

$$\min_x \left\{\frac{1}{p_i}c^T x + \frac{p_i}{p_i}.\min_y \left\{q(\xi^i)^T y : W(\xi^i)y = h(\xi^i) - T(\xi^i)x,\ y \geq 0\right\}\right\}$$
$$\text{s.t. } Ax = b,\ x \geq 0.$$

Now we have normed the probability of the second stage again, but we have also normed the first stage costs, so they are computed with weights relevant to the second stage costs. This problem replaces now the problem (2.36). It is clear that $\frac{1}{p_i}$ can be factored out before the first minimization and so we have to solve

$$\frac{1}{p_i}.\min_x \left\{c^T x + p_i.\min_y \left\{q(\xi^i)^T y : W(\xi^i)y = h(\xi^i) - T(\xi^i)x,\ y \geq 0\right\}\right\} \tag{2.41}$$
$$\text{s.t. } Ax = b,\ x \geq 0,$$

which is equivalent to solving

$$\begin{aligned}&\min_x \ \left\{c^\top x + p_i.\min_y \left\{q(\xi^i)^\top y : W(\xi^i)y = h(\xi^i) - T(\xi^i)x,\ y \geq 0\right\}\right\}\\&\text{s.t. } Ax = b,\ x \geq 0.\end{aligned} \tag{2.42}$$

The optimal solutions of the problems (2.41) and (2.42) are identical and the optimal objective function value is not of interest now.
Let us denote $x^*_{\{i\}}$ the optimal first stage solution of the last problem. We again define

$$MWS_{\{i\}}=\mathrm{E}_\xi z(x^*_{\{i\}},\xi)=c^\top x^*_{\{i\}}+\sum_{k=1}^{K} p_k.\min_y\left\{q(\xi^k)^\top y : W(\xi^k)y=h(\xi^k)-T(\xi^k)x^*_{\{i\}}, y\geq 0\right\}.$$

This definition is the same as that in (2.37) but we have obtained $x^*_{\{i\}}$ in a different way than before.
Again, $x^*_{\{i\}}$ is feasible for the recourse problem (2.35) and so $MWS_{\{i\}} \geq RP$. Hence,

$$U_1 := \min_{\{i\}\subseteq\{1,\dots,K\}} MWS_{\{i\}} \geq RP \geq WS =: L$$

which implies that

$$U_1 - L = \min_{\{i\}\subseteq\{1,\dots,K\}} MWS_{\{i\}} - WS = MEVPI_1 \geq EVPI = RP - WS,$$

which defines $MEVPI_1$.
Secondly, consider a subset S of the set $\{1,\dots,K\}$, S having N elements. We solve the problem for scenarios from the set S at first:

$$\begin{aligned}&\min_x \left\{\frac{1}{\sum_{l\in S} p^l} c^\top x + \sum_{n\in S} \frac{p^n}{\sum_{l\in S} p^l} \min_y \left\{q(\xi^n)^\top y : W(\xi^n)y = h(\xi^n) - T(\xi^n)x,\ y \geq 0\right\}\right\}\\&\text{s.t. } Ax = b,\ x \geq 0.\end{aligned}$$

Optimal first stage solution of this problem is the same as an optimal first stage solution of the problem

$$\begin{aligned}&\min_x \ \left\{c^\top x + \sum_{n\in S} p^n \min_y \left\{q(\xi^n)^\top y : W(\xi^n)y = h(\xi^n) - T(\xi^n)x,\ y \geq 0\right\}\right\}\\&\text{s.t. } Ax = b,\ x \geq 0,\end{aligned}$$

and we denote this optimal solution as x^*_S. Now we define MWS_S again:

$$MWS_S=\mathrm{E}_\xi z(x^*_S,\xi)=c^\top x^*_S+\sum_{k=1}^{K} p_k.\min_y\left\{q(\xi^k)^\top y : W(\xi^k)y=h(\xi^k)-T(\xi^k)x^*_S, y\geq 0\right\}.$$

x^*_S is feasible for the recourse problem (2.35) and so $MWS_S \geq RP$ and

$$U_N := \min_{\substack{S\subseteq\{1,\dots,K\}\\ \#S=N}} MWS_S \geq RP \geq WS =: L$$

which implies that

$$U_N - L = \min_{\substack{S\subseteq\{1,\dots,K\}\\ \#S=N}} MWS_S - WS = MEVPI_N \geq EVPI = RP - WS,$$

which redefines $MEVPI_N$.
We would like to decide whether $\min\limits_{\substack{S\subseteq\{1,\dots,K\}\\ \#S=N}} MWS_S \geq \min\limits_{\substack{Z\subseteq\{1,\dots,K\}\\ \#Z=N+1}} MWS_Z$ or vice versa. Again, there are two examples showing that none of the inequalities holds in general.

Example 2.44 When solving the problem for scenarios from the smaller set S, the decision maker hedges against N scenarios, while when solving the problem for the larger set Z, he hedges against $N+1$ scenarios, which, intuitively, should be better.
As an example, suppose the problem of $K = 3$ scenarios, probability of each scenario is $\frac{1}{3}$, $\#S = 1$, $\#Z = 2$, the only randomness is in the second stage costs:

$$q(\xi^1) = \begin{pmatrix} 1000 \\ 100 \\ 1 \end{pmatrix},\ q(\xi^2) = \begin{pmatrix} 1 \\ 1000 \\ 1 \end{pmatrix},\ q(\xi^2) = \begin{pmatrix} 1 \\ -3000 \\ 1 \end{pmatrix}; c = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}.$$

Recourse problem is formulated as follows:

$$\begin{aligned} \min_x \ & \left[\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}^T x + \frac{1}{3}\min_y \left\{ \begin{pmatrix} 1000 \\ 100 \\ 1 \end{pmatrix}^T y : y_i = 1 - x_i,\ i = 1,2,3 \right\} + \right. \\ & \left. + \frac{1}{3}\min_y \left\{ \begin{pmatrix} 1 \\ 1000 \\ 1 \end{pmatrix}^T y : y_i = 1 - x_i,\ i = 1,2,3 \right\} + \frac{1}{3}\min_y \left\{ \begin{pmatrix} 1 \\ -3000 \\ 1 \end{pmatrix}^T y : y_i = 1 - x_i,\ i = 1,2,3 \right\} \right] \\ \text{s.t. } & 0 \leq x_i \leq 1,\ i = 1,2,3. \end{aligned}$$

To compute $MWS_{\{1\}}$, we solve the problem

$$\begin{aligned} \min_x \ & \left[\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}^T x + \frac{1}{3}\min_y \left\{ \begin{pmatrix} 1000 \\ 100 \\ 1 \end{pmatrix}^T y : y_i = 1 - x_i,\ i = 1,2,3 \right\} \right] \\ \text{s.t. } & 0 \leq x_i \leq 1,\ i = 1,2,3, \end{aligned}$$

whose first stage optimal solution is $x^*_{\{1\}} = (1,1,0)^T$.

$$\begin{aligned} MWS_{\{1\}} = & \left[\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}^T \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \frac{1}{3}\min_y \left\{ \begin{pmatrix} 1000 \\ 100 \\ 1 \end{pmatrix}^T y : y_i = 1 - x^*_{\{1\}i},\ i = 1,2,3 \right\} + \right. \\ & + \frac{1}{3}\min_y \left\{ \begin{pmatrix} 1 \\ 1000 \\ 1 \end{pmatrix}^T y : y_i = 1 - x^*_{\{1\}i},\ i = 1,2,3 \right\} + \\ & \left. + \frac{1}{3}\min_y \left\{ \begin{pmatrix} 1 \\ -3000 \\ 1 \end{pmatrix}^T y : y_i = 1 - x^*_{\{1\}i},\ i = 1,2,3 \right\} \right] \\ = & \ 10 + 10 + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 21. \end{aligned}$$

Now we will compute $MWS_{\{2\}}$. The problem

$$\begin{aligned}\min_x \quad & \left[\begin{pmatrix}10\\10\\10\end{pmatrix}^T x + \frac{1}{3}\min_y\left\{\begin{pmatrix}1\\1000\\1\end{pmatrix}^T y : y_i = 1 - x_i,\ i = 1,2,3\right\}\right]\\ \text{s.t. } & 0 \le x_i \le 1,\ i = 1,2,3,\end{aligned}$$

has an optimal first stage solution $x^*_{\{2\}} = (0,1,0)^T$. We then have

$$\begin{aligned}MWS_{\{2\}} = & \left[\begin{pmatrix}10\\10\\10\end{pmatrix}^T \begin{pmatrix}0\\1\\0\end{pmatrix} + \frac{1}{3}\min_y\left\{\begin{pmatrix}1000\\100\\1\end{pmatrix}^T y : y_i = 1 - x^*_{\{2\}i},\ i = 1,2,3\right\} + \right.\\ & + \frac{1}{3}\min_y\left\{\begin{pmatrix}1\\1000\\1\end{pmatrix}^T y : y_i = 1 - x^*_{\{2\}i},\ i = 1,2,3\right\} +\\ & \left. + \frac{1}{3}\min_y\left\{\begin{pmatrix}1\\-3000\\1\end{pmatrix}^T y : y_i = 1 - x^*_{\{2\}i},\ i = 1,2,3\right\}\right]\\ = & \ 10 + \frac{1}{3}(1000+1) + \frac{1}{3}(1+1) + \frac{1}{3}(1+1) = 345.\end{aligned}$$

The first problem to be solved when computing $MWS_{\{3\}}$ reads

$$\begin{aligned}\min_x \quad & \left[\begin{pmatrix}10\\10\\10\end{pmatrix}^T x + \frac{1}{3}\min_y\left\{\begin{pmatrix}1\\-3000\\1\end{pmatrix}^T y : y_i = 1 - x_i,\ i = 1,2,3\right\}\right]\\ \text{s.t. } & 0 \le x_i \le 1,\ i = 1,2,3,\end{aligned}$$

and its optimal first stage solution is $x^*_{\{1\}} = (0,0,0)^T$. Then

$$\begin{aligned}MWS_{\{3\}} = & \left[\begin{pmatrix}10\\10\\10\end{pmatrix}^T \begin{pmatrix}0\\0\\0\end{pmatrix} + \frac{1}{3}\min_y\left\{\begin{pmatrix}1000\\100\\1\end{pmatrix}^T y : y_i = 1 - x^*_{\{3\}i},\ i = 1,2,3\right\} + \right.\\ & + \frac{1}{3}\min_y\left\{\begin{pmatrix}1\\1000\\1\end{pmatrix}^T y : y_i = 1 - x^*_{\{3\}i},\ i = 1,2,3\right\} +\\ & \left. + \frac{1}{3}\min_y\left\{\begin{pmatrix}1\\-3000\\1\end{pmatrix}^T y : y_i = 1 - x^*_{\{3\}i},\ i = 1,2,3\right\}\right]\\ = & \ 0 + \frac{1}{3}(1000+100+1) + \frac{1}{3}(1+1000+1) + \frac{1}{3}(1-3000+1) = -298\frac{1}{3}.\end{aligned}$$

Now we will compute $MWS_{\{1,2\}}$:

$$\begin{aligned}\min_x \quad & \left[\begin{pmatrix}10\\10\\10\end{pmatrix}^T x + \frac{1}{3}\min_y\left\{\begin{pmatrix}1000\\100\\1\end{pmatrix}^T y : y_i = 1 - x_i,\ i = 1,2,3\right\} + \right.\\ & \left. + \frac{1}{3}\min_y\left\{\begin{pmatrix}1\\1000\\1\end{pmatrix}^T y : y_i = 1 - x_i,\ i = 1,2,3\right\}\right]\\ \text{s.t. } & 0 \le x_i \le 1,\ i = 1,2,3,\end{aligned}$$

results in an optimal first stage solution $x^*_{\{1,2\}} = (1,1,0)^\top$ and

$$\begin{aligned} MWS_{\{1,2\}} = \Bigg[\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}^\top \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} &+ \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1000 \\ 100 \\ 1 \end{pmatrix}^\top y : y_i = 1 - x^*_{\{1,2\}i},\ i = 1,2,3 \right\} + \\ &+ \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1 \\ 1000 \\ 1 \end{pmatrix}^\top y : y_i = 1 - x^*_{\{1,2\}i},\ i = 1,2,3 \right\} + \\ &+ \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1 \\ -3000 \\ 1 \end{pmatrix}^\top y : y_i = 1 - x^*_{\{1,2\}i},\ i = 1,2,3 \right\} \Bigg] \\ = 10 + 10 + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 21. \end{aligned}$$

For computing $MWS_{\{1,3\}}$, the first problem to be solved is

$$\begin{aligned} \min_x \ \Bigg[\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}^\top x &+ \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1000 \\ 100 \\ 1 \end{pmatrix}^\top y : y_i = 1 - x_i,\ i = 1,2,3 \right\} + \\ &+ \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1 \\ -3000 \\ 1 \end{pmatrix}^\top y : y_i = 1 - x_i,\ i = 1,2,3 \right\} \Bigg] \\ \text{s.t. } 0 \le x_i \le 1,\ i = 1,2,3, \end{aligned}$$

and its optimal first stage solution is $x^*_{\{1,2\}} = (1,0,0)^\top$.

$$\begin{aligned} MWS_{\{1,3\}} = \Bigg[\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}^\top \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} &+ \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1000 \\ 100 \\ 1 \end{pmatrix}^\top y : y_i = 1 - x^*_{\{1,3\}i},\ i = 1,2,3 \right\} + \\ &+ \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1 \\ 1000 \\ 1 \end{pmatrix}^\top y : y_i = 1 - x^*_{\{1,3\}i},\ i = 1,2,3 \right\} + \\ &+ \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1 \\ -3000 \\ 1 \end{pmatrix}^\top y : y_i = 1 - x^*_{\{1,3\}i},\ i = 1,2,3 \right\} \Bigg] \\ = 10 + \frac{1}{3}(100+1) + \frac{1}{3}(1000+1) + \frac{1}{3}(-3000+1) = -622\frac{1}{3}. \end{aligned}$$

Finally, we will compute $MWS_{\{2,3\}}$:

$$\begin{aligned} \min_x \ \Bigg[\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}^\top x &+ \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1 \\ 1000 \\ 1 \end{pmatrix}^\top y : y_i = 1 - x_i,\ i = 1,2,3 \right\} + \\ &+ \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1 \\ -3000 \\ 1 \end{pmatrix}^\top y : y_i = 1 - x_i,\ i = 1,2,3 \right\} \Bigg] \\ \text{s.t. } 0 \le x_i \le 1,\ i = 1,2,3, \end{aligned}$$

has an optimal first stage solution $x^*_{\{2,3\}} = (0,0,0)^T$ and

$$\begin{aligned} MWS_{\{2,3\}} = \Bigg[& \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}^T \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1000 \\ 100 \\ 1 \end{pmatrix}^T y : y_i = 1 - x^*_{\{2,3\}i},\ i = 1,2,3 \right\} + \\ & + \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1 \\ 1000 \\ 1 \end{pmatrix}^T y : y_i = 1 - x^*_{\{2,3\}i},\ i = 1,2,3 \right\} + \\ & + \frac{1}{3} \min_y \left\{ \begin{pmatrix} 1 \\ -3000 \\ 1 \end{pmatrix}^T y : y_i = 1 - x^*_{\{2,3\}i},\ i = 1,2,3 \right\} \Bigg] \\ = 0 + \frac{1}{3}(1000 + 100 + 1) + \frac{1}{3}(1 + 1000 + 1) + \frac{1}{3}(1 - 3000 + 1) = -298\frac{1}{3}. \end{aligned}$$

Let's summarize our results. We obtained that

$$\begin{aligned} MWS_{\{1\}} &= 21, & MWS_{\{2\}} &= 345, & MWS_{\{3\}} &= -298\frac{1}{3}, \\ MWS_{\{1,2\}} &= 21, & MWS_{\{1,3\}} &= -622\frac{1}{3}, & MWS_{\{2,3\}} &= -298\frac{1}{3}. \end{aligned}$$

Thus we have the expected result

$$\min_{\{i\} \subseteq \{1,2,3\}} MWS_{\{i\}} > \min_{\substack{\{i,j\} \subseteq \{1,2,3\} \\ i \neq j}} MWS_{\{i,j\}}.$$

Example 2.45 It would be quite good to prove that the opposite inequality cannot work, but there is a counterexample again.

We have a problem with $K = 4$ scenarios, $\#S = 2$ and $\#Z = 3$. Probability of each scenario is $\frac{1}{4}$ and the only randomness enters in the second stage costs:

$$q(\xi^1) = \begin{pmatrix} 1000 \\ 0 \\ 1000 \\ 0 \\ -\frac{1}{3} \end{pmatrix},\ q(\xi^2) = \begin{pmatrix} 1000 \\ \frac{1}{1000} \\ 1000 \\ 0 \\ -\frac{1}{3} \end{pmatrix},\ q(\xi^3) = \begin{pmatrix} 0 \\ 1000 \\ 0 \\ 1000 \\ -\frac{1}{3} \end{pmatrix},\ q(\xi^3) = \begin{pmatrix} \frac{1}{1000} \\ 1000 \\ 0 \\ 1000 \\ -\frac{1}{3} \end{pmatrix};\ c = \begin{pmatrix} -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \end{pmatrix}$$

Constraints in the first stage are $0 \leq x_i \leq 1$ for $i = 1, \ldots, 4$ and constraints in the second stage are $y_i = x_i$ for $i = 1, \ldots, 4$, and $y_5 = \sum_{i=1}^{4} x_i$. For simplicity of writing, we denote $M(x) = \left\{ y \in \mathbb{R}^5 : y_i = x_i,\ i = 1, \ldots, 4,\ y_5 = \sum_{i=1}^{4} x_i \right\}$.

The full recourse problem reads

$$\min_x \left[\begin{pmatrix} -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \end{pmatrix}^T x + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 1000 \\ 0 \\ 1000 \\ 0 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 1000 \\ \frac{1}{1000} \\ 1000 \\ 0 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \right.$$

$$\left. + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 0 \\ 1000 \\ 0 \\ 1000 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} \frac{1}{1000} \\ 1000 \\ 0 \\ 1000 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} \right]$$

$$\text{s.t. } 0 \le x_i \le 1,\ i = 1, \ldots, 4.$$

At first, we will compute $MWS_{\{1,3\}}$:

$$\min_x \left[\begin{pmatrix} -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \end{pmatrix}^T x + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 1000 \\ 0 \\ 1000 \\ 0 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 0 \\ 1000 \\ 0 \\ 1000 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} \right]$$

$$\text{s.t. } 0 \le x_i \le 1,\ i = 1, \ldots, 4,$$

gives an optimal first stage solution $x^*_{\{1,3\}} = (0,0,0,0)^T$ which results in $MWS_{\{1,3\}} = 0$. Now we will compute $MWS_{\{1,2,3\}}$. At first we solve the problem

$$\min_x \left[\begin{pmatrix} -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \end{pmatrix}^T x + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 1000 \\ 0 \\ 1000 \\ 0 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 1000 \\ \frac{1}{1000} \\ 1000 \\ 0 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 0 \\ 1000 \\ 0 \\ 1000 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} \right]$$

$$\text{s.t. } 0 \le x_i \le 1,\ i = 1, \ldots, 4.$$

Its optimal first stage solution is $x^*_{\{1,3\}} = (0,1,0,1)^T$ and

$$\begin{aligned} MWS_{\{1,2,3\}} &= 2.\left(-249\frac{13}{16}\right) + \frac{1}{4}\left(0 - \frac{2}{3}\right) + \frac{1}{4}\left(\frac{1}{1000} - \frac{2}{3}\right) + \frac{1}{4}\left(1000 + 1000 - \frac{2}{3}\right) + \\ &\quad + \frac{1}{4}\left(1000 + 1000 - \frac{2}{3}\right) \\ &\doteq 499,71. \end{aligned}$$

To compute $MWS_{\{1,2,4\}}$, we solve

$$\min_x \left[\begin{pmatrix} -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \end{pmatrix}^T x + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 1000 \\ 0 \\ 1000 \\ 0 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 1000 \\ \frac{1}{1000} \\ 1000 \\ 0 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} \frac{1}{1000} \\ 1000 \\ 0 \\ 1000 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} \right]$$

$$\text{s.t. } 0 \le x_i \le 1,\ i = 1, \ldots, 4,$$

which gives an optimal first stage solution $x^*_{\{1,3\}} = (0,1,0,1)^T$. Then

$$\begin{aligned} MWS_{\{1,2,4\}} &= 2.\left(-249\frac{13}{16}\right) + \frac{1}{4}\left(0 - \frac{2}{3}\right) + \frac{1}{4}\left(\frac{1}{1000} - \frac{2}{3}\right) + \frac{1}{4}\left(1000 + 1000 - \frac{2}{3}\right) + \\ &\quad + \frac{1}{4}\left(1000 + 1000 - \frac{2}{3}\right) \\ &\doteq 499,71. \end{aligned}$$

The problem to be solved when computing $MWS_{\{1,3,4\}}$ reads

$$\min_x \left[\begin{pmatrix} -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \end{pmatrix}^T x + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 1000 \\ 0 \\ 1000 \\ 0 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 0 \\ 1000 \\ 0 \\ 1000 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} \frac{1}{1000} \\ 1000 \\ 0 \\ 1000 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} \right]$$
s.t. $0 \le x_i \le 1,\ i = 1, \ldots, 4.$

The first stage optimal solution is $x^*_{\{1,3\}} = (1,0,1,0)^T$ and

$$\begin{aligned} MWS_{\{1,3,4\}} &= 2.\left(-249\frac{13}{16}\right) + \frac{1}{4}\left(1000 + 1000 - \frac{2}{3}\right) + \frac{1}{4}\left(1000 + 1000 - \frac{2}{3}\right) + \\ &\quad + \frac{1}{4}\left(0 - \frac{2}{3}\right) + \frac{1}{4}\left(\frac{1}{1000} - \frac{2}{3}\right) \\ &\doteq 499,71. \end{aligned}$$

Finally, we will compute $MWS_{\{2,3,4\}}$:

$$\min_x \left[\begin{pmatrix} -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \\ -249\frac{13}{16} \end{pmatrix}^T x + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 1000 \\ \frac{1}{1000} \\ 1000 \\ 0 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} 0 \\ 1000 \\ 0 \\ 1000 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} + \frac{1}{4} \min_{y \in M(x)} \left\{ \begin{pmatrix} \frac{1}{1000} \\ 1000 \\ 0 \\ 1000 \\ -\frac{1}{3} \end{pmatrix}^T y \right\} \right]$$
s.t. $0 \le x_i \le 1,\ i = 1, \ldots, 4,$

results in $x^*_{\{1,3\}} = (1,0,1,0)^T$ and

$$\begin{aligned} MWS_{\{2,3,4\}} &= 2.\left(-249\frac{13}{16}\right) + \frac{1}{4}\left(1000 + 1000 - \frac{2}{3}\right) + \frac{1}{4}\left(1000 + 1000 - \frac{2}{3}\right) + \\ &\quad + \frac{1}{4}\left(0 - \frac{2}{3}\right) + \frac{1}{4}\left(\frac{1}{1000} - \frac{2}{3}\right) \\ &\doteq 499,71. \end{aligned}$$

To summarize this, we obtained $MWS_{\{1,3\}} = 0$ and so $\min\limits_{\substack{\{i,j\} \subseteq \{1,\ldots,4\} \\ i \neq j}} MWS_{\{i,j\}} \le 0$, while

$$\min_{\substack{\{i,j,k\} \subseteq \{1,\ldots,4\} \\ i \neq j \neq k \neq i}} MWS_{\{i,j,k\}} = MWS_{\{1,2,3\}} = MWS_{\{1,2,4\}} = MWS_{\{1,3,4\}} = MWS_{\{2,3,4\}} \doteq 499.71.$$

It is very easy to see that in fact only two scenarios are in this example, since the first scenario is almost the same as the second scenario and the same holds true for the third and fourth one. So this example is quite artificial and can serve just as a mathematical counterexample showing that the relevant inequality does not hold true. The idea of this example is that this (in fact, two-point with equal probabilities) distribution is very well represented by a pair of scenarios $\{1,3\}$ or $\{1,4\}$ or $\{2,3\}$ or $\{2,4\}$, but it is very distorted when any triplet of scenarios is taken into account.

Via counterexamples, we have just shown that no chain of inequalities works between $\min_{S_1} MWS_{S_1}, \min_{S_2} MWS_{S_2}, \ldots, \min_{S_K} MWS_{S_K}$ with $\#S_1 < \#S_2 < \ldots < \#S_K$. As a direct consequence, there is either no chain of inequalities between $MEVPI_1, \ldots, MEVPI_K$.

2.7. Sample information

In this section we will deal with a little more complicated problem presented in [13]. We consider a stochastic problem with a random variable ξ, but the distribution of ξ is not known exactly. It depends on a parameter $\theta \in \Theta$ which is also a random variable for us. We can obtain some information on the true value of θ by sampling. Then we can hedge against the future development of ξ more precisely, which leads to a "better" expected optimal objective function value. Therefore, we can derive value of sample information, which is not value of knowing the future development of ξ, but value of knowing (more) exactly the distribution of ξ.

Consider a two-stage stochastic recourse problem with right hand side only:

$$\begin{aligned} &\min_x \mathrm{E}_\xi z(x,\xi) = \min_x \left\{ c^T x + \mathrm{E}_\xi \min_y \{q^T y : Tx + Wy = \xi, y \geq 0\} \right\} \\ &\text{s.t.} \ \ x \in K, \end{aligned} \tag{2.43}$$

where K is a closed convex polyhedron, ξ is an r-component real random vector ($r \in \mathbb{N}$) defined on a probability space $(\Omega, \mathscr{A}, P)$ with a convex support Ξ and a finite expectation. Vector q and matrices T and W are now non-random and $h = \xi$. We assume an existence of all densities which are needed in this chapter.
Deterministic equivalent to the stochastic problem (2.43) reads

$$\begin{aligned} &\min_x \mathrm{E}_\xi z(x,\xi) = \min_x \left\{ c^T x + \mathrm{E}_\xi Q(x,\xi) \right\} \\ &\text{s.t.} \ \ x \in K \end{aligned} \tag{2.44}$$

where

$$Q(x,\xi) = \min_y \{q^T y : Tx + Wy = \xi, y \geq 0\} \text{ for all } (x,\xi) \in K \times \Xi.$$

Suppose that $F(t,\theta) = P(\xi \leq t|\theta)$ is a distribution function of ξ, depending on a parameter θ. This θ is a k-component vector of unknown parameters, $\theta \in \Theta \subseteq \mathbb{R}^k$. A prior distribution function of the random variable θ is denoted as $G(.)$ and $G(.|\xi)$ is a posterior distribution function of θ given ξ.
Let S be an m-component vector of sufficient statistics for the family $\{F(t,\theta), \theta \in \Theta\}$.
For any given value $\tilde{\theta}$ of θ, the independent identically distributed random variables $\xi^1, \ldots, \xi^n$ form a random sample of size n from the distribution $F(t,\tilde{\theta})$, and $S_n(\xi^1, \ldots, \xi^n)$ is the corresponding sufficient statistics; the dimension m of S (or S_n) is fixed for all n

and the subscript n denotes the size of the sample used for computing the value of S_n. Denote as $w_n(s|\tilde{\theta})$ the conditional density of $S_n(.)$ given $\theta = \tilde{\theta}$.
The posterior distribution function of θ given $S_n(\xi^1, \ldots, \xi^n) = s$ is

$$G_n(u|s) = P(\theta \leq u | S_n(\xi^1, \ldots, \xi^n) = s) = \frac{\int_{-\infty}^{u} w_n(s|\theta) G(d\theta)}{\int w_n(s|t) G(dt)}.$$

Note that in both of these integrals we integrate over k-dimensional sets. The first one is $(-\infty; u_1\rangle \times \ldots \times (-\infty; u_k\rangle$ and the second one is $\mathbb{R}^k$. In this chapter, in all the integrals without bounds explicitly written, we integrate over the whole space of an appropriate dimension (r or k).
Let $H_n(t|s) = P(\xi \leq t | S_n = s) = \int F(t, \theta) G_n(d\theta|s)$ be a predictive distribution function of ξ given $S_n(\xi^1, \ldots, \xi^n) = s$. We write $H_0(t) = P(\xi \leq t)$ in the case of no sampling. As is shown in [5], there exists a family Γ of distributions G with the property that if the prior (no sample) distribution G of θ belongs to Γ, then for all $n \in \mathbb{N}$ and for all $S_n(\xi^1, \ldots, \xi^n) = s$, the posterior (conditional) distribution $G(.|s)$ belongs to Γ as well.
The decision problem without observations (with prior distribution $G(.)$ of θ) then reads

$$\begin{aligned} \min_{x \in K} \mathrm{E}_\xi z_0(x, \xi | G) &= \min_{x \in K} \left\{ c^T x + \int Q(x, t) H_0(dt) \right\} \\ &= \min_{x \in K} \left\{ c^T x + \iint Q(x, t) F(dt, \theta) G(d\theta) \right\}. \end{aligned} \tag{2.45}$$

The only purpose of the subscript 0 in z_0 is to emphasize that we deal with a problem with no sampling; in fact, there is $z_0 = z$.
Denote as $R(G)$ the optimal objective function value of the problem (2.45) and as x_0^* its optimal solution. So there is

$$R(G) = \min_{x \in K} \mathrm{E}_\xi z_0(x, \xi | G) = \mathrm{E}_\xi z_0(x_0^*, \xi | G)$$

which represents the risk (created by the uncertain future realization of ξ and by its unknown distribution) for the case of no observations.

Suppose that the assumptions 2.1 and 2.2 are satisfied. Then a general theorem can be formulated as a standard result of the duality theory:
Theorem 2.46 For the problem (2.44) it holds

$$Q(x, \xi) = \max_u \{ (\xi - Tx)^T u : W^T u \leq q^T \}.$$

□

Theorem 2.47 Suppose that Γ is a convex set. Then $R(G)$ is concave in G on Γ.
Proof. If $G^1, G^2 \in \Gamma$ and $\alpha \in \langle 0, 1\rangle$ then $\alpha G^1 + (1-\alpha) G^2 \in \Gamma$ represents the compound of the two distributions G^1 and G^2 with probabilities α and $1-\alpha$, respectively. Function $\hat{z}_0$: $\hat{z}_0(x, G) = \mathrm{E}_\xi z_0(x, \xi | G)$ is linear in G, since integral is a linear functional. Hence,

$$\begin{aligned} R(\alpha G^1 + (1-\alpha) G^2) &= \min_{x \in K} \hat{z}_0(x, \alpha G^1 + (1-\alpha) G^2) = \min_{x \in K} \left(\alpha \hat{z}_0(x, G^1) + (1-\alpha) \hat{z}_0(x, G^2) \right) \\ &\geq \alpha \min_{x \in K} \hat{z}_0(x, G^1) + (1-\alpha) \min_{x \in K} \hat{z}_0(x, G^2) = \alpha R(G^1) + (1-\alpha) R(G^2). \end{aligned}$$

□

Theorem 2.48 If the assumptions 2.1 and 2.2 hold, then the decision problem (2.45) without observations is a convex program.

Proof. The problem reads $\min_{x \in K} c^T x + \mathrm{E}_\xi Q(x, \xi)$. The first stage part $c^T x$ is linear in x, so it is enough to show that $\mathrm{E}_\xi Q(x, \xi)$ is convex in x. Since the expectation E is a linear functional, it is sufficient to prove that $Q(x, \xi)$ is convex in x. According to the theorem 2.46, we have that

$$Q(x, \xi) = \max_u \{(\xi - Tx)^T u : W^T u \leq q^T\} = \max_{u \in U} f_\xi(u, x),$$

where we have denoted $f_\xi(u, x) = (\xi - Tx)^T u$ and $U = \{u : W^T u \leq q^T\}$. U is a convex set and f_ξ is linear in x for all $u \in U$. Hence, for every $\alpha \in (0, 1)$ and for every $x^1, x^2 \in K$ it holds

$$\begin{aligned} Q(\alpha x^1 + (1-\alpha)x^2, \xi) &= \max_{u \in U} f_\xi(u, \alpha x^1 + (1-\alpha)x^2) = \max_{u \in U} \left(\alpha f_\xi(u, x^1) + (1-\alpha) f_\xi(u, x^2)\right) \\ &\leq \alpha \max_{u \in U} f_\xi(u, x^1) + (1-\alpha) \max_{u \in U} f_\xi(u, x^2) \\ &= \alpha Q(x^1, \xi) + (1-\alpha) Q(x^2, \xi). \end{aligned}$$

□

Given a sample $(\xi^1, \ldots, \xi^n)$ of size n with $S_n(\xi^1, \ldots, \xi^n) = s$, the deterministic equivalent of the original problem (2.43) is

$$\begin{aligned} \min_{x \in K} \mathrm{E}_{\xi|s} z_n(x, \xi | G_n(.|s)) &= \min_{x \in K} \left\{ c^T x + \int Q(x, t) H_n(dt|s) \right\} \\ &= \min_{x \in K} \left\{ c^T x + \iint Q(x, t) F(dt, \theta) G_n(d\theta|s) \right\}. \end{aligned}$$

If $x_n^*(s)$ is an optimal solution of this problem, then the posterior risk is

$$z_n^*(G_n(.|s)) := \min_{x \in K} \mathrm{E}_{\xi|s} z_n(x, \xi | G_n(.|s)) = \mathrm{E}_{\xi|s} z_n(x_n^*(s), \xi | G_n(.|s)),$$

so $x_n^*(s)$ is a decision function (of s) for the distribution $G_n(.|s)$ for $n = 1, 2, \ldots$

We can formulate another linear problem, corresponding to an expected value problem given $S_n(\xi^1, \ldots, \xi^n) = s$:

$$\begin{aligned} \min_{x,y} \ & \bar{z}_n(x, y|s) = c^T x + q^T y \\ \text{s.t. } & x \in K, \\ & Tx + Wy = \mu_n(s), \\ & y \geq 0, \end{aligned}$$

where $\mu_n(s) = \mathrm{E}(\xi | S_n = s) = \int t H_n(dt|s)$ is a conditional expectation of ξ given $S_n = s$. Denote the optimal objective function value of this problem as $\bar{z}_n^*(s)$. Then the following holds:

Theorem 2.49 Given a sample of size n, $S_n(\xi^1, \ldots, \xi^n) = s$, then

(a) there is $z_n^*(G_n(.|s)) \geq \bar{z}_n^*(s)$,

(b) $z_n^*(G_n(.|s))$ is concave in G_n on Γ.

Proof. The proof can be found in [13]. □

Remark 2.50 The inequality in (a) corresponds to inequality which can be intuitively written as

$$(RP \text{ given } S_n = s) \geq (EV \text{ given } S_n = s).$$

Define

$$R_n(G) = \mathrm{E}_s[z_n^*(G_n(.|s))]$$

the expected value (w.r.t. s) of the posterior risk given a sample of size n, such that $S_n(\xi^1, \ldots, \xi^n) = s$. Then $R_n(G)$ has the following properties:

Theorem 2.51
(a) $R_n(G)$ is concave in G on Γ.
(b) $R_n(G)$ is a monotonic nonincreasing function of n.
Proof. The proof can be found in [13]. □

Now we can define *expected value of sample information* (of sample of size n) as

$$EVSI_n = \mathrm{E}_\xi z_0(x_0^*, \xi|G) - \mathrm{E}_s[z_n^*(G_n(.|s))] = R(G) - R_n(G). \tag{2.46}$$

Theorem 2.52
(a) $EVSI_n$ is a monotonic nondecreasing and nonnegative function of the size n of a sample.
(b) $EVSI_n \leq \mathrm{E}_\xi z_0(x_0^*, \xi|G) - \mathrm{E}_s \bar{z}_n^*(s)$, i.e. $EVSI_n \leq R(G) - \mathrm{E}_s \bar{z}_n^*(s)$.
Proof.
(a) Follows immediately from the theorem 2.51 (b).
(b) Follows immediately from definitions and theorem 2.49 (a). □

Assume that we have the possibility to increase the sample size n up to infinity. Then we can define the limiting posterior distribution function of θ given $S_\infty(\xi^1, \xi^2, \ldots) = s$. It is assumed to be a one-point distribution of the form

$$G_\infty(u) = \begin{cases} 0 \text{ if } u < \theta \\ 1 \text{ if } u \geq \theta, \end{cases}$$

which represents full knowledge of the true value of θ. The posterior risk with a such an infinite sampling is

$$R_\infty(G) = \int \min_{x \in K} \left[c^T x + \int Q(x,t) F(dt, \theta) \right] G(d\theta)$$

and $R_n(G) \searrow R_\infty(G)$ as $n \to \infty$, since $R_n(G)$ is a monotonous nonincreasing function of n. *Expected value of infinite sample information* is then defined as

$$EVSI_\infty = R(G) - R_\infty(G)$$

and it is easy to see that $EVSI_n \nearrow EVSI_\infty$ as $n \to \infty$.

Note that $EVSI_\infty$ is not the value of knowing future realizations of ξ, but the value of knowing the distribution of ξ precisely and having the possibility to solve the standard stochastic problem.

2.8. Chance-constrained problems

We will touch the problem of chance-constrained programs only in brief because they are at the edge of the family of stochastic programs we are interested in.

2.8.1. Value of information in chance-constrained programming

Consider a simple chance-constrained problem of the form

$$
\begin{aligned}
RP = \min_{x} \ & z(x) \\
\text{s.t. } & x \in K \subseteq \mathbb{R}, \\
& P(x \geq \xi) \geq \alpha,
\end{aligned} \tag{2.47}
$$

where z is a real convex function bounded on K, K is an interval, ξ is a real scalar random variable with known distribution, and $\alpha \in (0;1)$ is a given real constant. The meaning of this problem is that we have to make a decision x which is greater or equal to ξ with a given probability α. We can consider (2.47) as a partially relaxed constraint $x \geq \xi$ a.s. Of course, no realization of ξ is observed in the time of making the decision.

Denote $L = \{x : P(x \geq \xi) \geq \alpha\}$. We will assume that the set $K \cap L$ is nonempty and closed and that $\min_{x \in K \cap L} z(x)$ exists (finite).

We can equivalently reformulate our problem as

$$
\begin{aligned}
RP = \min_{(x(\xi),\hat{x})} \ & \left[z(x(\xi)) + \tilde{I}_{[P(x(\xi) \geq \xi) \geq \alpha]}(x(\xi)) \right] \\
\text{s.t. } & x(\xi) \in K \text{ a.s.}, \\
& x(\xi) = \hat{x} \text{ a.s.}
\end{aligned} \tag{2.48}
$$

where the penalization function $\tilde{I}$ is defined as

$$
\tilde{I}_{[A(x)]}(x) = \begin{cases} 0 & \text{when the statement } A(x) \text{ is true for the argument } x \\ +\infty & \text{otherwise.} \end{cases}
$$

The objective function in (2.48) is deterministic almost surely and so an expectation E_ξ is not needed.

The related wait-and-see problem is formulated analogically, only the nonanticipativity constraints are relaxed so that the optimal solution is a function of ξ, and the expectation E_ξ is needed:

$$
\begin{aligned}
WS = \mathrm{E}_\xi \min_{x(\xi)} \ & \left[z(x(\xi)) + \tilde{I}_{[P(x(\xi) \geq \xi) \geq \alpha]}(x(\xi)) \right] \\
\text{s.t. } \ & x(\xi) \in K \text{ a.s.}
\end{aligned} \tag{2.49}
$$

Generalization for several probabilistic constraints or for joint chance-constraints or combination of those is obvious. In both (2.48) and (2.49) the statement $A(x)$ which stands as a parameter of the penalization function $\tilde{I}$ would change only.

Remark 2.53 We are aware of the unpleasant fact that $\tilde{I}$ is an extended real-valued function. We assume that $-\infty < z(x) < +\infty$ for all $x \in K$, so the value of the objective function $z + \tilde{I}$ is bounded (when the value of $\tilde{I}$ is 0) or it is $+\infty$. It holds $\{x : \tilde{I}_{[P(x \geq \xi) \geq \alpha]}(x) = 0\} = L \neq \emptyset$ and so the minimizing function $x(\xi)$ in (2.49) will be such that $x(\tilde{\xi}) \in L$ for all realizations $\tilde{\xi}$ of ξ. This assures that the expectation in (2.49) exists.

Theorem 2.54 Consider the chance-constrained program (2.48) and its wait-and-see modification (2.49). Define *value of prefect information for the chance-constrained problem* (2.48) as

$$
EVPI^C = RP - WS.
$$

Then $EVPI^C$ is nonnegative.

Proof. It is obvious that $WS = \mathrm{E}_\xi g(\xi)$ for $g(\xi) = \min\limits_{x(\xi)\in K} \left[z(x(\xi)) + \tilde{I}_{[P(x(\xi)\geq \xi)\geq \alpha]}(x(\xi)) \right]$ and $g(\xi) \leq RP$ a.s., which implies that $WS \leq RP$. □

Remark 2.55 A notion of value of partial information for a special case of a chance-constrained problem can be found in [14]. The authors, however, introduced a definition of value of partial information which we do not consider as absolutely correct, and they have also shown that "their" value of partial information can be negative.

2.8.2. Sample information in chance-constrained programming

In this part, we return to the problem introduced in chapter 2.7. and we employ a very similar notation. We deal with a problem with a simple linear objective function and probability constraints. The constraints depend on a realization of a random variable ξ which is not known in the time of making the decision, and in addition the distribution of this random variable is not known exactly—it is considered as a function of a parameter θ which is seen as another random variable. Its true value can be detected by sampling to some extent. The question is, what is the worth of information gained by the sampling. We will deal with problems with multiple and joint probabilistic constraints. We will omit some proofs and details; they can be found in [13].

Consider a simple problem with a linear deterministic objective function and probability constraint of the form

$$\begin{aligned} &\min_{x\in K} z(x) = c^T x \\ &\text{s.t. } P\left(\xi_i \leq \sum\nolimits_j a_{ij}x_j\right) \leq \alpha_i,\ i = 1,\dots,r. \end{aligned} \tag{2.50}$$

This constraint is equivalent with

$$F_i\left(\sum\nolimits_j a_{ij}x_j, \theta_i\right) \leq \alpha_i,\ i = 1,\dots,r,\ \text{for any true value of } \theta_i,$$

where $F_i(t,\theta_i) = P(\xi_i \leq t|\theta_i)$. Now, $\theta_i = (\theta_{i1},\dots,\theta_{im_i})$ is an m_i-component vector of unknown parameters and for every i, θ_i has a prior (no sample) distribution function $G_i(.)$. We consider fixed values $\alpha_1,\dots,\alpha_r$.

Let $(\xi_i^1,\dots,\xi_i^n)$ be random sample of size n from the distribution $F_i(.,\theta_i)$, $\theta_i \in \Theta_i$, and consider a sufficient statistics $S_n(\xi_i^1,\dots,\xi_i^n)$ for the family $F_i(.,\theta_i)$, $\theta_i \in \Theta_i$, for $i = 1,\dots,r$. Denote $H_{i,n}(t,s^i)$, $i = 1,\dots,r$, the posterior distribution function of ξ_i given $S_n(\xi_i^1,\dots,\xi_i^n) = s^i$. Again, $H_{i,0} = H_{i,0}(t)$ corresponds to the case of no sampling. Naturally, we could generalize this problem for samples of various sizes $n_1,\dots,n_r$, but it would not be too interesting while complications in writing would be great.

Denote $F = (F_1,\dots,F_r)$, let $b = b(z,F_1,\dots,F_r) = b(z,F)$ be a badness or negative utility function, continuous and monotonic increasing in $(z,F) \in \mathbb{R} \times \langle 0,1\rangle^r$. Define minimum of expected negative-utility of costs in the case of no sampling as

$$\Phi(G) = \min_{x\in K} \mathrm{E}_\theta b\left(c^T x, F_1\left(\sum\nolimits_j a_{1j}x_j,\theta_1\right),\dots,F_r\left(\sum\nolimits_j a_{rj}x_j,\theta_r\right)\right) \tag{2.51}$$

where G is a prior distribution function of $\theta = (\theta_1,\dots,\theta_r)$. (Now, θ is a vector of vectors.) Thanks to the negative-utility function b we can penalise violating of any probability

constraints, putting e.g. $b(c^T x, F_1(\sum_j a_{1j}x_j, \theta_1), \ldots, F_r(\sum_j a_{rj}x_j, \theta_r)) = +\infty$ if there exists $i \in \{1, \ldots, K\}$ such that $F_i(\sum_j a_{ij}x_j, \theta_i) > \alpha_i$ for all $x \in K$.
We can define a problem which is analogical to (2.51) but is constructed after samples of size n are made:

$$\Phi\left(G_n(.|s)\right) = \min_{x \in K} \mathrm{E}_{G_n} b\left(c^T x, F_1\left(\sum\nolimits_j a_{1j}x_j, \theta_1\right), \ldots, F_r\left(\sum\nolimits_j a_{rj}x_j, \theta_r\right)\right)$$

where E_{G_n} is a symbolical denotation for $\mathrm{E}_{\theta|s}$ with $s = (s^1, \ldots, s^r)$, $s^i = S_n(\xi_i^1, \ldots, \xi_i^n)$, $i = 1, \ldots, r$.
We can define *expected value of sample information* of the sample of size n for the problem (2.51) as

$$EVSI_n^C = \Phi(G) - \mathrm{E}_s \Phi(G_n(.|s)) \tag{2.52}$$

where $s = (s^1, \ldots, s^r)$, $s^i = S_n(\xi_i^1, \ldots, \xi_i^n)$, $i = 1, \ldots, r$. (The upper index C stands for Chance-constraint.)
Some properties analogical to those for $EVSI_n$ can be derived: $EVSI_n^C \geq 0$ for all distributions G, for all $n \in \mathbb{N}$, and it is a nondecreasing function of $n \geq 1$.
For an increasing n, limit posterior distribution of θ_i given $S_\infty(\xi_i^1, \xi_i^2, \ldots) = s^i$ for $i = 1, \ldots, r$ and $s = (s^1, \ldots, s^r)$, is $G_{i,\infty}(.)$. It is assumed to be a one-point distribution of the form

$$G_{i,\infty}(t_i) = \begin{cases} 0 & \text{if} \quad t_i < \theta_i \\ 1 & \text{if} \quad t_i \geq \theta_i, \ i = 1, \ldots, r. \end{cases}$$

Denote $G_\infty = (G_{1,\infty}, \ldots, G_{r,\infty})$. Analogically as in (2.52), we can define $EVSI_\infty^C$ such that $EVSI_n^C \nearrow EVSI_\infty^C$ for $n \to \infty$.

Consider a problem similar to (2.50) but now the probability constraints have to be satisfied for all $i = 1, \ldots, r$ simultaneously:

$$\begin{aligned} &\min_{x \in K} c^T x \\ &\text{s.t. } P\left(\xi_i \leq \sum\nolimits_j a_{ij}x_j, \ i = 1, \ldots, r\right) \leq \alpha \end{aligned} \tag{2.53}$$

for a given $\alpha \in \langle 0, 1\rangle$. Let $F(t, \theta) = P(\xi_i \leq t_i, \ i = 1, \ldots, r|\theta)$ be the joint distribution function of $\xi = (\xi_1, \ldots, \xi_r)$ depending on $\theta \in \Theta$.
Let again $b = b(c^T x, F)$ be a negative-utility function continuous and monotonic increasing in $(c^T x, F) \in \mathbb{R} \times \langle 0, 1\rangle$. We can define negative utility of the costs in the case of no observation as

$$\Phi^{JC}(G) = \min_{x \in K} \mathrm{E}_\theta b\left(c^T x, F\left(\sum\nolimits_j a_{1j}x_j, \ldots, \sum\nolimits_j a_{rj}x_j, \theta\right)\right).$$

Having observed $\xi^1, \ldots, \xi^n$, we can define

$$EVSI_n^{JC} = \Phi^{JC}(G) - \mathrm{E}_s \Phi^{JC}(G_n(.|s)),$$

where $G_n(.|s)$ is a joint posterior distribution function of θ given $S_n(\xi^1, \ldots, \xi^n) = s$ and where the expectation is taken w.r.t. the value s of the sufficient statistics S_n. Then again, $EVSI_n^{JC} \geq 0$ for all distribution functions G and for all $n \in \mathbb{N}$, and $EVSI_n^{JC}$ is a monotonic nondecreasing function of $n \geq 1$.

Assume again that an infinite sampling $\xi^1, \xi^2, \ldots$ is available. Then we can compute the posterior distribution function of ξ given θ (which is given by the sampling) and define $EVSI_\infty^{JC}$ in an obvious way. Then again $EVSI_n^{JC} \nearrow EVSI_\infty^{JC}$ for $n \to \infty$.

2.9. Value of perfect information as a risk measure.

In some applications, the notion of a *risk measure* is used for a functional with some properties. In [1], a definition and some interesting properties of *coherent risk measures* are introduced. For us, risk measure will be a functional defined on some space of random variables with values on the real line. We will show, following there partly the approach presented in [15] and in [16], that in some cases, value of information can be used as a risk measure. We will also compare properties of the value of information with properties of coherent risk measures given in [1].

2.9.1. Special case with a piecewise linear objective function

Let's have a probability space $(\Omega, \tilde{\mathscr{F}}, P)$ and a random variable $\xi\colon \Omega \to \mathbb{R}$ which represents outcomes. We will again deal with an optimization problem related to this random variable. The decision maker has to decide about an income x which has to be available after the realization of the random outcome or consumption ξ is observed (but the decision is done before the realization of ξ). The goal is to get just the outcome as precisely as possible; any possible surplus will be discounted by $d \in (0;1)$, as the surplus is less satisfactory than consumption; any shortfall will bring additional costs so that shortfall total costs are q per unit, $q > 1$ (d, q given). For any σ-field $\mathscr{F} \subseteq \tilde{\mathscr{F}}$ denote as $\mathscr{M}(\mathscr{F}, \Omega, \mathbb{R})$ the space of all $\mathscr{F}$-measurable functions $f\colon \Omega \to \mathbb{R}$.
In this paragraph we consider a problem with a simple special case of the objective function:

$$\min_x \mathrm{E}\left(q[\xi - x]^+ - d[\xi - x]^- + x\right)$$

and we will add some constraints.

Suppose for a while that the decision maker knows the future realization of the random variable ξ. Then he solves a deterministic optimization problem and his solution is $x = x(\xi)$, that is, x is a function of ξ, we have that $(x(\xi))(\omega) = x(\xi(\omega)),\ x\colon \Omega \to \mathbb{R}$ and it is an $\tilde{\mathscr{F}}$-measurable mapping. Hence, the anticipative problem under full information on the future development of ξ reads

$$\begin{aligned} &\min_x \mathrm{E}\left(q[\xi - x]^+ - d[\xi - x]^- + x\right) \\ &\text{s.t. } x \in \mathscr{M}(\tilde{\mathscr{F}}, \Omega, \mathbb{R}). \end{aligned}$$

The optimal objective function value of this problem (the *WS* value) is denoted as $V^\xi_{\tilde{\mathscr{F}}}$, the optimal solution is $x(\xi) = \xi$ a.s., i.e., x is just the value which is to be realized. The optimal objective function value is $\mathrm{E}(q.0 - d.0 + \xi) = \mathrm{E}\xi$.
Suppose now that the decision maker solves the problem under a partial (imperfect) information on the future development of ξ. His decision variable is now a function $x\colon \Omega \to \mathbb{R}$ which is $\mathscr{F}$-measurable for some σ-field $\mathscr{F} \subseteq \tilde{\mathscr{F}}$. The larger this σ-field is, the better informed is the decision maker. In this situation we will say that information (represented by the σ-field) $\mathscr{F}$ is available. The decision problem is then

$$\begin{aligned} &\min_x \mathrm{E}\left(q[\xi - x]^+ - d[\xi - x]^- + x\right) \\ &\text{s.t. } x \in \mathscr{M}(\mathscr{F}, \Omega, \mathbb{R}) \end{aligned} \tag{2.54}$$

and the optimal objective function value is denoted as $V^{\xi}_{\mathscr{F}}$.

Theorem 2.56 For two σ-fields $\mathscr{F}_1,\ \mathscr{F}_2$ such that $\mathscr{F}_1 \subseteq \mathscr{F}_2$, there is $V^{\xi}_{\mathscr{F}_1} \geq V^{\xi}_{\mathscr{F}_2}$.

Proof. If $\mathscr{F}_1 \subseteq \mathscr{F}_2$, then $x^{-1}(A) \in \mathscr{F}_1$ implies that $x^{-1}(A) \in \mathscr{F}_2$ for every function x and every set $A \in B(\mathbb{R})$. Hence, the feasibility set for $V^{\xi}_{\mathscr{F}_2}$ is larger or equal to the feasibility set for $V^{\xi}_{\mathscr{F}_1}$, and so $V^{\xi}_{\mathscr{F}_1} \geq V^{\xi}_{\mathscr{F}_2}$. □

The risk contained in the random variable ξ and the incomplete information represented by the σ-field $\mathscr{F}$ is the difference between the clairvoyant's minimal costs (computed with full information $\tilde{\mathscr{F}}$) and the human decision-maker's minimal costs (computed with incomplete information $\mathscr{F}$):

$$R^{\xi}_{\mathscr{F}} = V^{\xi}_{\mathscr{F}} - V^{\xi}_{\tilde{\mathscr{F}}} = V^{\xi}_{\mathscr{F}} - \mathrm{E}[\xi],$$

which defines a *risk value of perfect information* $R^{\xi}_{\mathscr{F}}$.

It is clear that $R^{\xi}_{\mathscr{F}} \geq 0$, since $V^{\xi}_{\mathscr{F}} \geq V^{\xi}_{\tilde{\mathscr{F}}}$.

The *expected value of partial information* given by the σ-field $\mathscr{F}$ is $V^{\xi}_{\mathscr{F}_0} - V^{\xi}_{\mathscr{F}}$, where $\mathscr{F}_0 = \{\emptyset, \Omega\}$ is the simplest σ-field, and it is nonnegative.

Definition 2.57 Let ξ^1 and ξ^2 be two random outcome variables.

We say that *first order stochastic dominance* holds, if $\mathrm{E}[f(\xi^1)] \leq \mathrm{E}[f(\xi^2)]$ for all nondecreasing functions f for which these expected values are finite. We write $\xi^1 \prec_{FSD} \xi^2$.

We say that *second order stochastic dominance* holds, if $\mathrm{E}[f(\xi^1)] \leq \mathrm{E}[f(\xi^2)]$ for all nondecreasing and convex functions f for which these expected values are finite. We write $\xi^1 \prec_{SSD} \xi^2$.

We say that *convex dominance* holds, if $\mathrm{E}[f(\xi^1)] \leq \mathrm{E}[f(\xi^2)]$ for all convex functions f for which these expected values are finite. We write $\xi^1 \prec_{CC} \xi^2$.

Definition 2.58 If $G_i,\ i = 1,\dots,k$ are distribution functions and $p_i,\ i = 1,\dots,k$ are probabilities such that $\sum_{i=1}^{k} p_i = 1$, then *compound distribution* of G_i with probabilities p_i, $i = 1,\dots,k$, is defined as

$$G(u) = \sum_{i=1}^{k} p_i G_i(u).$$

It is the distribution of the *compound random variable* C of random variables $W_1,\dots,W_k$ with distribution functions $G_1,\dots,G_k$, respectively, with probabilities $p_1,\dots,p_k$, defined as

$$C = \mathscr{C}(W_1,\dots,W_k; p_1,\dots,p_k) = \begin{cases} W_1 \text{ if } I = 1 \\ W_2 \text{ if } I = 2 \\ \dots \\ W_k \text{ if } I = k \end{cases}$$

where I is a random variable, which is independent on $W_1,\dots,W_k$ and which satisfies $P(I = i) = p_i$ for $i = 1,\dots,k$.

Definition 2.59 The probability functional F is *compound convex (concave)*, if it satisfies for all G_i and p_i from the preceding definition

$$\mathrm{F}\left[\sum_{i=1}^{k} p_i G_i\right] \leq\ (\geq)\ \sum_{i=1}^{k} p_i \mathrm{F}\left[G_i\right].$$

The probability functional F is *compound linear*, if equality holds there.

Let us consider the simplest σ-field $\mathscr{F}_0 = \{\emptyset, \Omega\}$. Note that x is $\mathscr{F}_0$-measurable, if it is a (real) constant. Then

$$V_{\mathscr{F}_0}^{\xi} = \min_x \mathrm{E}[z(x,\xi)]$$
$$\text{s.t. } x \in \mathscr{M}(\mathscr{F}_0, \Omega, \mathbb{R}),$$

where

$$\begin{aligned} z(x,\xi) &= q[\xi - x]^+ + x - d[\xi - x]^- \\ &= q[\xi - x]^+ + d(\xi - x) - d[\xi - x]^+ + x \\ &= (q-d)[\xi - x]^+ + (1-d)x + d\xi. \end{aligned}$$

Function $z = z(x,\xi)$ is nondecreasing and convex in ξ and so $\xi^1 \geq \xi^2$ a.s. implies that $z(x,\xi^1) \geq z(x,\xi^2)\ \forall x$ a.s. which further implies that $\min_x \mathrm{E}z(x,\xi^1) \geq \min_x \mathrm{E}z(x,\xi^2)$, that is $V_{\mathscr{F}}^{\xi^1} \geq V_{\mathscr{F}}^{\xi^2}$ which means that higher outcomes generate higher costs.
It also holds that $\xi^1 \prec_{SSD} \xi^2$ implies $V_{\mathscr{F}_0}^{\xi^1} \leq V_{\mathscr{F}_0}^{\xi^2}$, $\xi^1 \prec_{FSD} \xi^2$ implies that $V_{\mathscr{F}_0}^{\xi^1} \leq V_{\mathscr{F}_0}^{\xi^2}$ and $\xi^1 \prec_{CC} \xi^2$ implies that $V_{\mathscr{F}_0}^{\xi^1} \leq V_{\mathscr{F}_0}^{\xi^2}$. It can also be derived that $\xi^1 \prec_{CC} \xi^2$ implies that $R_{\mathscr{F}_0}^{\xi^1} \leq R_{\mathscr{F}_0}^{\xi^2}$: The function $z(x,\xi)$ is convex in ξ for all x and so is $z(x,\xi) - \xi$. Hence, $\xi_1 \prec_{CC} \xi_2$ implies by definition that

$$\mathrm{E}[z(x,\xi_1) - \xi_1] \leq \mathrm{E}[z(x,\xi_2) - \xi_2] \quad \forall x \in \mathbb{R}$$

and so

$$\min_x \mathrm{E}[z(x,\xi_1) - \xi_1] = \min_x \mathrm{E}[z(x,\xi_1)] - \mathrm{E}[\xi_1] \leq \min_x \mathrm{E}[z(x,\xi_2) - \xi_2] = \min_x \mathrm{E}[z(x,\xi_2)] - \mathrm{E}[\xi_2]$$

which means that $R_{\mathscr{F}_0}^{\xi_1} \leq R_{\mathscr{F}_0}^{\xi_2}$. The same does not hold for $\mathscr{F} \neq \mathscr{F}_0$, as is stated in [16]. The reason is obvious: We know that z is a convex function of ξ for all $x \in \mathbb{R}$, but x can somehow depend on ξ for $x \in \mathscr{M}(\mathscr{F}, \Omega, \mathbb{R})$ and the form of this dependence is unknown. Then it is not sure that z is convex in ξ.
For every σ-field $\mathscr{F} \subseteq \tilde{\mathscr{F}}$ and for all $b \in \mathbb{R}$ it holds

$$\begin{aligned} V_{\mathscr{F}}^{\xi+b} &= \min_x \mathrm{E}(q[\xi + b - x]^+ - d[\xi + b - x]^- + x) \\ &= \min_{x-b} \mathrm{E}(q[\xi - (x-b)]^+ - d[\xi - (x-b)]^- + (x-b)) + b \\ &= V_{\mathscr{F}}^{\xi} + b \end{aligned}$$

which means that increasing outcome by a known (constant) amount b leads to an increase of overall costs by the same amount. It can be simply seen now that $R_{\mathscr{F}}^{\xi}$ is translation-invariant, i.e. $R_{\mathscr{F}}^{\xi+b} = R_{\mathscr{F}}^{\xi}$ for all b real constants.
Also, $V_{\mathscr{F}}, R_{\mathscr{F}}$ are positively homogeneous, i.e. for any $\lambda \geq 0$ it holds

$$V_{\mathscr{F}}^{\lambda\xi} = \min_{\lambda x} \mathrm{E}z(\lambda x, \lambda\xi) = \min_{\lambda x} \mathrm{E}\lambda.z(x,\xi) = \lambda \min_x \mathrm{E}z(x,\xi) = \lambda V_{\mathscr{F}}^{\xi},$$

because

$$z(\lambda x, \lambda\xi) = q[\lambda\xi - \lambda x]^+ - d[\lambda\xi - \lambda x]^- + \lambda x = \lambda(q[\xi - x]^+ - d[\xi - x]^- + x) = \lambda z(x,\xi).$$

$R^{(.)}_{\mathscr{F}}$ is a linear combination of two positively homogeneous functions, so it is also positively homogeneous.
Note that $z = z(x, \xi)$ is convex in (x, ξ). Let ξ, ζ be two random outcome variables defined on $(\Omega, \tilde{\mathscr{F}}, P)$ (dependence is possible) and $V^{\xi}_{\mathscr{F}} = \mathrm{E}z(x_1^*, \xi)$ and $V^{\zeta}_{\mathscr{F}} = \mathrm{E}z(x_2^*, \zeta)$. Then for any fixed constant $p \in (0,1)$ it holds

$$\begin{aligned} V^{p\xi+(1-p)\zeta}_{\mathscr{F}} &= \min_x \mathrm{E}z(x, p\xi + (1-p)\zeta) \\ &\leq \mathrm{E}z(px_1^* + (1-p)x_2^*, p\xi + (1-p)\zeta) \\ &\leq \mathrm{E}\left[pz(x_1^*, \xi) + (1-p)z(x_2^*, \zeta)\right] \text{ since } z \text{ is convex in } (x, \xi) \\ &= pV^{\xi}_{\mathscr{F}} + (1-p)V^{\zeta}_{\mathscr{F}}. \end{aligned}$$

So $V^{\xi}_{\mathscr{F}}$ is convex in ξ (and $V^{\xi}_{\tilde{\mathscr{F}}}$ is linear in ξ, since $V^{\xi}_{\tilde{\mathscr{F}}} = \mathrm{E}\xi$) and so $R^{\xi}_{\mathscr{F}}$ is also convex in ξ. This means that combining two outcomes cannot increase risk.
Consider two random income variables ξ and ζ and let

$$\theta = \begin{cases} \xi & \text{with probability } p \\ \zeta & \text{with probability } 1-p, \end{cases}$$

where $p \in (0,1)$. Then for any σ-field $\mathscr{F} \subseteq \tilde{\mathscr{F}}$ is

$$\mathrm{E}z(x, \theta) = p\mathrm{E}z(x, \xi) + (1-p)\mathrm{E}z(x, \zeta) \quad \forall x \in \mathscr{M}(\mathscr{F}, \Omega, \mathbb{R})$$

and so

$$\min_x \mathrm{E}z(x, \theta) \geq \min_x p\mathrm{E}z(x, \xi) + \min_x (1-p)\mathrm{E}z(x, \zeta)$$

which means that $V^{\theta}_{\mathscr{F}} \geq pV^{\xi}_{\mathscr{F}} + (1-p)V^{\zeta}_{\mathscr{F}}$. Together with the obvious equality $V^{\theta}_{\tilde{\mathscr{F}}} = pV^{\xi}_{\tilde{\mathscr{F}}} + (1-p)V^{\zeta}_{\tilde{\mathscr{F}}}$ this implies that

$$\begin{aligned} R^{\theta}_{\mathscr{F}} = V^{\theta}_{\mathscr{F}} - V^{\theta}_{\tilde{\mathscr{F}}} &\geq \left[pV^{\xi}_{\mathscr{F}} + (1-p)V^{\zeta}_{\mathscr{F}}\right] - \left[pV^{\xi}_{\tilde{\mathscr{F}}} + (1-p)V^{\zeta}_{\tilde{\mathscr{F}}}\right] \\ &= p\left[V^{\xi}_{\mathscr{F}} - V^{\xi}_{\tilde{\mathscr{F}}}\right] + (1-p)\left[V^{\zeta}_{\mathscr{F}} - V^{\zeta}_{\tilde{\mathscr{F}}}\right] \\ &= pR^{\xi}_{\mathscr{F}} + (1-p)R^{\zeta}_{\mathscr{F}}, \end{aligned}$$

i.e., $R^{(.)}_{\mathscr{F}}$ is compound concave.
Define

$$R^{\xi}_{\mathscr{F}_1,\mathscr{F}_2} = V^{\xi}_{\mathscr{F}_1} - V^{\xi}_{\mathscr{F}_2}$$

for $\mathscr{F}_1 \subseteq \mathscr{F}_2$ some σ-subfields of $\tilde{\mathscr{F}}$. Hence, $R^{\xi}_{\mathscr{F}_1,\mathscr{F}_2}$ is a difference between the optimal objective function values of a less (with $\mathscr{F}_1$) and more (with $\mathscr{F}_2$) informed decision maker and it can be seen as a *value of increasing information* from $\mathscr{F}_1$ to $\mathscr{F}_2$. It holds

$$R^{\xi}_{\mathscr{F}_1,\mathscr{F}_2} = (V^{\xi}_{\mathscr{F}_1} - V^{\xi}_{\tilde{\mathscr{F}}}) - (V^{\xi}_{\mathscr{F}_2} - V^{\xi}_{\tilde{\mathscr{F}}}) = R^{\xi}_{\mathscr{F}_1} - R^{\xi}_{\mathscr{F}_2},$$

so $R^{\xi}_{\mathscr{F}_1,\mathscr{F}_2}$ corresponds to a decrease of risk when the available information increases from $\mathscr{F}_1$ to (better) $\mathscr{F}_2$. Some trivial properties can be derived: $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \mathscr{F}_3 \subseteq \tilde{\mathscr{F}} \Rightarrow R^{\xi}_{\mathscr{F}_1,\mathscr{F}_3} \geq R^{\xi}_{\mathscr{F}_1,\mathscr{F}_2}$, and $b \in \mathbb{R} \Rightarrow R^{\xi+b}_{\mathscr{F}_1,\mathscr{F}_2} = R^{\xi}_{\mathscr{F}_1,\mathscr{F}_2}$, and $R^{\xi}_{\mathscr{F},\tilde{\mathscr{F}}} = R^{\xi}_{\mathscr{F}} \quad \forall \mathscr{F} \subseteq \tilde{\mathscr{F}}$.

An extreme case is the difference between the clairvoyant and the totally uninformed decision maker: $R^{\xi}_{\mathscr{F}_0,\tilde{\mathscr{F}}} = R^{\xi}_{\mathscr{F}_0} - 0 = V^{\xi}_{\mathscr{F}_0} - V^{\xi}_{\tilde{\mathscr{F}}}$ which represents an expected value of perfect information.

It was argued that $V^{\xi}_{\mathscr{F}}$, $R^{\xi}_{\mathscr{F}}$ should be used as risk measures. Properties required to be satisfied by *coherent risk measures* are formulated in [1], and we can compare the properties of $V^{\xi}_{\mathscr{F}}$, $R^{\xi}_{\mathscr{F}}$ with them. Reformulated a little in an appropriate way, the properties of a coherent risk measure ϱ are:
(1) *Transition equivariance:* $\varrho(X+\alpha) = \varrho(X) + \alpha$ for all $\alpha \in \mathbb{R}$, X random variable.
This property is completely satisfied by $V^{\xi}_{\mathscr{F}}$ and it is not satisfied by $R^{\xi}_{\mathscr{F}}$.
(2) *Subaditivity:* $\varrho(X_1 + X_2) \leq \varrho(X_1) + \varrho(X_2)$ for all X_1, X_2 random variables.
We have proven that $R^{\xi}_{\mathscr{F}}$ is convex in ξ, i.e. $R^{p\xi+(1-p)\zeta}_{\mathscr{F}} \leq p.R^{\xi}_{\mathscr{F}} + (1-p).R^{\zeta}_{\mathscr{F}}$ for all $\xi, \zeta \in (\Omega, \tilde{\mathscr{F}}, P)$, $p \in (0,1)$. We can write $X_1 = \frac{1}{2}\xi$ and $X_2 = \frac{1}{2}\zeta$. Then there is

$$\begin{aligned} R^{X_1+X_2}_{\mathscr{F}} \leq \frac{1}{2}R^{\xi}_{\mathscr{F}} + \frac{1}{2}R^{\zeta}_{\mathscr{F}} &= \frac{1}{2}R^{2X_1}_{\mathscr{F}} + \frac{1}{2}R^{2X_2}_{\mathscr{F}} \\ &= \frac{1}{2}.2R^{X_1}_{\mathscr{F}} + \frac{1}{2}.2R^{X_2}_{\mathscr{F}} \text{ since } R^{(\cdot)}_{\mathscr{F}} \text{ si positively homogenous} \\ &= R^{X_1}_{\mathscr{F}} + R^{X_2}_{\mathscr{F}}. \end{aligned}$$

So this property is fully satisfied by $R^{\xi}_{\mathscr{F}}$. Exactly the same is true for $V^{\xi}_{\mathscr{F}}$.
(3) *Positive homogeneity:* $\varrho(tX) = t.\varrho(X)$ for all $t \in \mathbb{R}^+$ and X random variable.
This is satisfied for both $V^{\xi}_{\mathscr{F}}$ and $R^{\xi}_{\mathscr{F}}$.
(4) *Monotonicity:* for all $X, Y \in (\Omega, \tilde{\mathscr{F}}, P)$ with $X \leq Y$ a.s. it holds $\varrho(X) \leq \varrho(Y)$.
It is satisfied for $V^{\xi}_{\mathscr{F}}$ but it is not satisfied for $R^{\xi}_{\mathscr{F}}$.
(5) *Relevance:* $\forall X \in (\Omega, \tilde{\mathscr{F}}, P)$, $X \geq 0$ a.s., X is not a constant a.s., we have $\varrho(X) > 0$.
This additional property ensures us that we do not consider a trivial case.
A positive outcome must always imply positive overall costs, hence $V^{\xi}_{\mathscr{F}} > 0$. We have also proved that $R^{\xi}_{\mathscr{F}} \geq 0$, but there can exist a σ-field $\mathscr{F} \subset \tilde{\mathscr{F}}$ such that any σ-field $\tilde{\tilde{\mathscr{F}}} \supseteq \mathscr{F}$ gives an information sufficient for eliminating the risk $R^{\xi}_{\tilde{\tilde{\mathscr{F}}}}$ to zero.

Our conclusion is that the optimal objective function value $V^{\xi}_{\mathscr{F}}$ of the problem (2.54) is a coherent risk measure in the sense of [1], while $R^{\xi}_{\mathscr{F}}$ is not. The reasons are translation equivariance, monotonicity and relevance not being satisfied by $R^{\xi}_{\mathscr{F}}$.

Remark 2.60 In economics, the *conditional value at risk* (*CVaR*) is often used to judge a risk brought by an uncertain investment. Relationships between $R^{\xi}_{\mathscr{F}}$ and *CVaR* are derived in [16].

2.9.2. General objective function

Let ξ be a random outcome variable defined on some probability space $(\Omega, \tilde{\mathscr{F}}, P)$, x an amount to be available after the realization of ξ, but to be decided about before this realization. The goal of the decision maker is to find

$$\min_{x} \mathrm{F} z(x, \xi),$$

where F is a probability functional (in the preceding part, we had $\mathrm{F} \equiv \mathrm{E}$), i.e. F is a mapping from $(\Omega, \tilde{\mathscr{F}}, P)$ to $\mathbb{R}$; we will assume that F is monotonic w.r.t. stochastic dominance of the first and second order and to the concave dominance. This means that for ϕ^1, ϕ^2 random variables, $\phi^1 \prec_{FSD} \phi^2$ or $\phi^1 \prec_{SSD} \phi^2$ or $\phi^1 \prec_{CC} \phi^2$ imply that $\mathrm{F}(\phi^1) \leq \mathrm{F}(\phi^2)$. The function $z = z(x, \xi)$ is a measurable real-valued objective (outcome) function of (x, ξ) which is convex and nondecreasing in ξ. Again, for a σ-field $\mathscr{F} \subseteq \tilde{\mathscr{F}}$ is $\mathscr{M}(\mathscr{F}, \Omega, \mathbb{R})$ the space of all $\mathscr{F}$-measurable functions $x\colon \Omega \to \mathbb{R}$. We assume that all the employed minima exist.

The nonanticipative optimization problem (analogical to the here-and-now problem) is then

$$\min_x \mathrm{F} z(x, \xi)$$
$$\text{s.t. } x \in \mathscr{M}(\mathscr{F}, \Omega, \mathbb{R}),$$

where again $\mathscr{F}$ represents available information, $\mathscr{F} \subseteq \tilde{\mathscr{F}}$. Let us denote the optimal objective function value of this problem as $V^{\xi}_{\mathscr{F}}$. Again, it is clear that $\mathscr{F}_1 \subseteq \mathscr{F}_2$ implies that $V^{\xi}_{\mathscr{F}_1} \geq V^{\xi}_{\mathscr{F}_2}$.

The anticipative clairvoyant's problem reads

$$\min_x \mathrm{F} z(x, \xi)$$
$$\text{s.t. } x \in \mathscr{M}(\tilde{\mathscr{F}}, \Omega, \mathbb{R})$$

and its optimal objective function value is $V^{\xi}_{\tilde{\mathscr{F}}}$.

The risk contained in the random variable ξ and the imperfect information $\mathscr{F}$ is again $R^{\xi}_{\mathscr{F}} = V^{\xi}_{\mathscr{F}} - V^{\xi}_{\tilde{\mathscr{F}}}$ and it also holds true that $R^{\xi}_{\mathscr{F}} \geq 0$, since $V^{\xi}_{\mathscr{F}} \geq V^{\xi}_{\tilde{\mathscr{F}}}$ for any σ-field $\mathscr{F} \subseteq \tilde{\mathscr{F}}$.

The "expected" value of partial information given by the σ-field $\mathscr{F}$ is $V^{\xi}_{\mathscr{F}_0} - V^{\xi}_{\mathscr{F}} \geq 0$, where $\mathscr{F}_0 = \{\emptyset, \Omega\}$ is the simplest σ-field, relating to a totally uninformed decision maker. Now the word "expected" is not used in a mathematical sense because a general functional F is used instead of E.

The function z is nondecreasing and convex in ξ and F is monotonic w.r.t. first order stochastic dominance. Hence, $\xi^1 \geq \xi^2$ a.s. implies that $V^{\xi^1}_{\mathscr{F}} \geq V^{\xi^2}_{\mathscr{F}}$ for all $\mathscr{F} \subseteq \tilde{\mathscr{F}}$ which means that higher outcome generates higher total costs.

It also holds that $\xi^1 \prec_{SSD} \xi^2$ implies $V^{\xi^1}_{\mathscr{F}_0} \leq V^{\xi^2}_{\mathscr{F}_0}$, $\xi^1 \prec_{FSD} \xi^2$ implies that $V^{\xi^1}_{\mathscr{F}_0} \leq V^{\xi^2}_{\mathscr{F}_0}$ and $\xi^1 \prec_{CC} \xi^2$ implies that $V^{\xi^1}_{\mathscr{F}_0} \leq V^{\xi^2}_{\mathscr{F}_0}$.

It is not generally true now that $\xi^1 \prec_{CC} \xi^2$ implies $R^{\xi_1}_{\mathscr{F}_0} \leq R^{\xi^2}_{\mathscr{F}_0}$, since $R^{\xi}_{\mathscr{F}_0} = V^{\xi}_{\mathscr{F}_0} - V^{\xi}_{\tilde{\mathscr{F}}}$ where $V^{\xi}_{\mathscr{F}_0}$ is monotonic w.r.t. convex dominance while $-V^{\xi}_{\tilde{\mathscr{F}}}$ is not.

Let us now assume that F is a convex functional and that z is convex in (x, ξ). Let $p \in (0, 1)$ be a real constant, let ξ, ζ be two random outcome variables (dependence is possible) and $V^{\xi}_{\mathscr{F}} = \min_x \mathrm{F}(z(x, \xi)) = \mathrm{F}(z(x^*_1, \xi))$, $V^{\zeta}_{\mathscr{F}} = \min_x \mathrm{F}(z(x, \zeta)) = \mathrm{F}(z(x^*_2, \zeta))$.

Then

$$\begin{aligned}
V_{\mathscr{F}}^{p\xi+(1-p)\zeta} &= \min_x \mathrm{F}(z(x, p\xi + (1-p)\zeta)) \\
&\le \mathrm{F}(z(px_1^* + (1-p)x_2^*, p\xi + (1-p)\zeta)) \\
&\le \mathrm{F}(pz(x_1^*, \xi) + (1-p)z(x_2^*, \zeta)) \ \text{ since } z \text{ is convex in } (x, \xi) \\
&\le p.\mathrm{F}(z(x_1^*, \xi)) + (1-p).\mathrm{F}(z(x_2^*, \zeta)) \ \text{ since F is convex} \\
&= p.V_{\mathscr{F}}^{\xi} + (1-p).V_{\mathscr{F}}^{\zeta},
\end{aligned}$$

so $V_{\mathscr{F}}^{(\cdot)}$ is convex in ξ.

This does not hold in general for $R_{\mathscr{F}}^{(\cdot)}$ since this is a difference of two convex functions.

It can also be easily derived that for F compound concave also $V_{\mathscr{F}}^{(\cdot)}$ is compound concave.

If we assume that, in addition, $V_{\tilde{\mathscr{F}}}^{(\cdot)}$ is compound linear, then also $R_{\mathscr{F}}^{(\cdot)}$ is compound concave.

We can again compare properties of the generalized $V_{\mathscr{F}}^{\xi}$ and $R_{\mathscr{F}}^{\xi}$ with the properties of coherent risk measures given in [1]. The result is quite poor: The properties are generally not satisfied by both $V_{\mathscr{F}}^{\xi}$ and $R_{\mathscr{F}}^{\xi}$, except for the monotonicity which is satisfied for $V_{\mathscr{F}}^{\xi}$ only. So $V_{\mathscr{F}}^{\xi}$ and $R_{\mathscr{F}}^{\xi}$ are not coherent risk measures in this generalized case.

At the end, let us define $R_{\mathscr{F}_1,\mathscr{F}_2}^{\xi} = V_{\mathscr{F}_1}^{\xi} - V_{\mathscr{F}_2}^{\xi} = R_{\mathscr{F}_1}^{\xi} - R_{\mathscr{F}_2}^{\xi}$ for $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \tilde{\mathscr{F}}$ two σ-fields. Again, $R_{\mathscr{F}_1,\mathscr{F}_2}^{\xi}$ expresses a decrease of risk when the available information becomes $\mathscr{F}_2$ instead of $\mathscr{F}_1$.

It is clear that $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \mathscr{F}_3$ implies $R_{\mathscr{F}_1,\mathscr{F}_3}^{\xi} \ge R_{\mathscr{F}_1,\mathscr{F}_2}^{\xi}$ and $R_{\mathscr{F},\tilde{\mathscr{F}}}^{\xi} = R_{\mathscr{F}}^{\xi}$ for all $\mathscr{F} \subseteq \tilde{\mathscr{F}}$. The translation-invariance property does not hold any more.

The extreme case is the difference between a totally uninformed decision maker and a clairvoyant (having full information): $R_{\mathscr{F}_0,\tilde{\mathscr{F}}}^{\xi} = V_{\mathscr{F}_0}^{\xi} - V_{\tilde{\mathscr{F}}}^{\xi}$ and this is “expected” value of perfect information. We cannot call this value expected value of perfect information, because we use a general probability functional F instead of expectation.

Remark 2.61 In [16] and [15] similar notions are defined for a special multiperiod problem. We will not return to them in the chapter devoted to multiperiod problems because they are not too interesting when we focus on various types of value of information.

Remark 2.62 The authors of [17] introduced an approach to the value of information based on optimization of convex risk measures (i.e. extended real-valued functionals; they call them *risk functions* and assume their convexity, monotonicity and translation equivariance). In their framework, the optimization problem reads

$$\inf_{x \in K} \varrho(F(x))$$

where ϱ is a risk measure and F is a function which maps the decision x to a random variable $\xi = F(x)$. After employing sophisticated tools based on dual spaces and functional analysis theorems, the authors show that for certain spaces of decisions x and random variables $\xi = F(x)$ it holds

$$\inf_{x \in K} \varrho(F(x)) \ge \varrho\left(\inf_{x \in K} F(x)\right)$$

where they define

$$\left(\inf_{x\in K} F(x)\right)(\omega) := \inf_{x\in K} \{(F(x))(\omega)\}.$$

This allows them to define *risk value of perfect information* (connected with the risk measure ϱ) as

$$RVPI_{\varrho} = \inf_{x\in K} \varrho(F(x)) - \varrho\left(\inf_{x\in K} F(x)\right) \geq 0$$

and *expected value of perfect information* (connected with a probability measure μ) as

$$EVPI_{\mu} = \inf_{x\in K} \mathrm{E}_{\mu}[F(x)] - \mathrm{E}_{\mu}\left[\inf_{x\in K} F(x)\right].$$

It is finally shown that

$$\inf_{\mu\in\mathscr{A}_{\mathscr{P}}(\varrho)} EVPI_{\mu} \leq RVPI_{\varrho} \leq \sup_{\mu\in\mathscr{A}_{\mathscr{P}}(\varrho)} EVPI_{\mu}$$

where $\mathscr{A}_{\mathscr{P}}(\varrho)$ is a set of all probability measures μ such that $\langle\mu;\xi\rangle := \int_{\Omega} \xi(\omega)d\mu(\omega) \leq \varrho(\xi)$ for all random variables ξ from the considered space.
This approach can be used to formulate a precise analogy of a so called interchangeability theorem (commutativity of infimum and expectation in recourse problems is proven under some specific conditions) introduced in [19]. It continues with dualisation of explicit nonanticipativity constraints which leads to consideration of *EVPI* as an "information price system"—the whole amount saved up by being able to find out the realization of random variable ξ would be paid for an inquiry necessary for getting this piece of information. This fact is derived independently in [19] and in [17] in different symbols but as a precise parallel. For details, see the above mentioned articles.

3. Multistage problems

3.1. General formulation of multistage problems

We often need to solve problems covering more then two time periods and more then two decision points. Formulation of such problems is of interest in the *multistage programming.*

Assumptions given here represent a natural generalization of the assumptions given for the two-stage problems. Suppose that the number of time periods is $T,\ T \geq 2$. Suppose for simplicity that there is one decision point at the beginning of each time period. (This need not be true in real situations, because sometimes we have to make a decision for more then one time period, or make several decisions in one period). Again, randomness is brought to the problem by a vector of real random variables $\xi_1, \dots, \xi_{T-1}$. For every $t = 1, \dots, T-1,\ \xi_t$ is defined on a probability space $(\Omega_t, \mathscr{F}_t, P)$. Probability distribution of all random variables is known with distribution functions $F(\xi_t)$ which are independent of the decisions $x_1, \dots, x_T$. Realization of the component ξ_t is observed in the t-th time period after making the decision x_t.

The order of making decisions and observing realizations of the random variables is $x_1, \xi_1, x_2, \xi_2, \dots, \xi_{T-1}, x_T$. Sometimes we use for technical purposes a component ξ_0 before x_1; this component ξ_0 may be considered as random, equal to a given constant with probability one. We will write $\xi^t = (\xi_1, \dots, \xi_t)$ and $x^t = (x_1, \dots, x_t)$.

The decision process is nonanticipative, which means that a decision x_t is allowed to depend on former decisions $x_1, \dots, x_{t-1}$ and former realizations $\xi_1, \dots, \xi_{t-1}$ which are already known (observed) in the time of making the decision x_t, but x_t has to be independent of any future realizations of random variables and future decisions.

Structure of the general T-stage problem is quite unfriendly and it is quite unpleasant to deal with it. It can be written as T nested problems as follows:

$$
\begin{aligned}
RP = \min_{x_1} \mathrm{E}_{\xi_1} f_0(x_1, \xi_1) &= \min_{x_1} \left\{ f_{10}(x_1) + \mathrm{E}_{\xi_1}[\varphi_1(x_1, \xi_1)] \right\} \\
&\text{s.t. } f_{1i}(x_1) \leq 0,\ i = 1, \dots, m_1,\ x_1 \in X_1 \\
\text{for } \varphi_1(x_1, \xi_1) &= \min_{x_2} \left\{ f_{20}(x_2) + \mathrm{E}_{\xi_2|\xi_1}[\varphi_2(x_1, x_2, \xi_1, \xi_2)] \right\} \\
&\quad \text{s.t. } f_{2i}(x_1, x_2, \xi_1) \leq 0,\ i = 1, \dots, m_2,\ x_2 \in X_2 \\
&\dots \\
\text{for } \varphi_{t-1}(x_1, \dots, x_{t-1}, \xi_1, \dots, \xi_{t-1}) &= \min_{x_t} \left\{ f_{t0}(x_t) + \mathrm{E}_{\xi_t|\xi^{t-1}}[\varphi_t(x_1, \dots, x_t, \xi_1, \dots, \xi_t)] \right\} \\
&\quad \text{s.t. } f_{ti}(x_1, \dots, x_t, \xi_1, \dots, \xi_{t-1}) \leq 0, i = 1, \dots, m_t, \\
&\quad\quad x_t \in X_t, \\
&\dots \\
\text{for } \varphi_T = 0,&
\end{aligned}
$$

where for all t and i are f_{ti} and φ_t real functions, $X_t \subseteq \mathbb{R}$ are nonempty sets and we deal with conditional expectations. In this form, we have divided the non-random and the random part of the objective function in each stage to create the best possible analogy to the two-stage programs. The resulting optimal objective function value is denoted as RP.

We will usually write the problem in a simpler form:

$$\min_{x_1 \in X_1} \mathrm{E}_{\xi_1} \min_{x_2 \in X_2(x_1,\xi_1)} \mathrm{E}_{\xi_2|\xi_1} \dots \mathrm{E}_{\xi_{T-1}|\xi^{T-2}} \min_{x_T \in X_T(x^{T-1},\xi^{T-1})} z(x_1,\dots,x_T,\xi_1,\dots,\xi_{T-1})$$

for an objective function z.
Now, the feasibility sets for the second and any further minimization have to be written in the form showing their dependence on former decisions and realizations, e.g. $X_2(x_1, \xi_1)$. The form of this dependence can be clearly seen in the following linear problem.
If all the objective functions and constraints are linear, we can write the T-stage problem as follows:

$$\begin{aligned} RP = \min_{x_1} \Big[c_1^T x_1 + \mathrm{E}_{\xi_1} \min_{x_2} \Big[c_2^T x_2 + \mathrm{E}_{\xi_2|\xi_1} \min_{x_3} \Big[c_3^T x_3 + \dots + \mathrm{E}_{\xi_{T-1}|\xi^{T-2}} \min_{x_T} \big[c_T^T x_T \big] \dots \Big]\Big]\Big] \\ \text{s.t. } & T_1 x_1 = h_1, \\ & T_2(\xi_1)x_1 + W_2(\xi_1)x_2 = h_2(\xi_1) \text{ a.s.}, \\ & T_3(\xi^2)x_2 + W_3(\xi^2)x_3 = h_3(\xi^2) \text{ a.s.}, \\ & \dots \\ & T_T(\xi^{T-1})x_{T-1} + W_T(\xi^{T-1})x_T = h_T(\xi^{T-1}) \text{ a.s.}, \\ & l_1 \le x_1 \le u_1, \\ & l_t(\xi^{t-1}) \le x_t \le u_t(\xi^{t-1}) \text{ a.s. for } t = 2,\dots,T. \end{aligned} \tag{3.1}$$

We can also rewrite the objective function in (3.1) as

$$\min_{x_1} \mathrm{E}_{\xi_1} \min_{x_2} \mathrm{E}_{\xi_2|\xi_1} \min_{x_3} \dots \mathrm{E}_{\xi_{T-1}|\xi^{T-2}} \min_{x_T} z(x_1,\dots,x_T,\xi_1,\dots,\xi_{T-1}),$$

where

$$z(x_1,\dots,x_T,\xi_1,\dots,\xi_{T-1}) = c_1^T x_1 + c_2^T x_2 + \dots + c_T^T x_T.$$

In general there is $c_t = c_t(\xi^{t-1})$ for $t = 2,\dots,T$.

It is much easier to solve the multistage problems when the distribution of the random variables $\xi_1,\dots,\xi_{T-1}$ is discrete. This means that for $t = 1,\dots,T-1$ the support of ξ_t is $\Xi_t = \{\xi_t^1,\dots,\xi_t^{S_t}\}$, S_t finite. We then speak about *scenario-based problems.*

3.2. Various approaches to the multistage problems

We have more possibilities of modelling a given multistage problem than we had for a two-stage one. In the following parts, we will formulate the problem (3.1) using several different approaches in order to be able to compare to each other their resulting optimal objective function values.

3.2.1. Wait-and-see problem

We can formulate the wait-and-see problem in a similar way as for the two-stage programs. In this case, the decision maker is allowed to postpone all his decisions to the end of the last time period, that is, until he knows all the realizations of the random variables, or he is able to reveal future realizations of all random variables at the beginning of the first time period. Thus all decisions $x_1,\dots,x_T$ are functions of ξ^{T-1}. The *wait-and-see*

T-stage linear problem reads

$$
\begin{aligned}
WS = \mathrm{E}_{\xi^{T-1}} \min_{x_1(\xi^{T-1}),\ldots,x_T(\xi^{T-1})} & \left\{ c_1^T x_1(\xi^{T-1}) + \ldots + c_T^T x_T(\xi^{T-1}) \right\} \\
\text{s.t. } & T_1 x_1(\xi^{T-1}) = h_1 \quad \text{a.s.}, \\
& T_2(\xi_1) x_1(\xi^{T-1}) + W_2(\xi_1) x_2(\xi^{T-1}) = h_2(\xi_1) \quad \text{a.s.}, \\
& T_3(\xi^2) x_2(\xi^{T-1}) + W_3(\xi^2) x_3(\xi^{T-1}) = h_3(\xi^2) \quad \text{a.s.}, \\
& \ldots \\
& T_T(\xi^{T-1}) x_{T-1}(\xi^{T-1}) + W_T(\xi^{T-1}) x_T(\xi^{T-1}) = h_T(\xi^{T-1}) \quad \text{a.s.}, \\
& l_1 \le x_1(\xi^{T-1}) \le u_1 \quad \text{a.s.}, \\
& l_t(\xi^{t-1}) \le x_t(\xi^{T-1}) \le u_t(\xi^{t-1}) \quad \text{a.s. for } t = 2, \ldots, T.
\end{aligned}
$$

This decision process is anticipative, since all the decisions can depend on all realizations, which are supposed to be known in advance.

3.2.2. Expected value problem

Now we replace the random variables $\xi_1, \ldots, \xi_{T-1}$ by their expectations $\bar{\xi}_1 = \mathrm{E}\xi_1, \ldots, \bar{\xi}_{T-1} = \mathrm{E}\xi_{T-1}$, respectively. We obtain a deterministic *expected value problem*

$$
\begin{aligned}
EV = \min_{x_1,\ldots,x_T} & \; z(x_1, \ldots, x_T, \bar{\xi}_1, \ldots, \bar{\xi}_{T-1}) \\
\text{s.t. } & T_1 x_1 = h_1, \\
& T_2(\bar{\xi}_1) x_1 + W_2(\bar{\xi}_1) x_2 = h_2(\bar{\xi}_2), \\
& \ldots \\
& T_T(\bar{\xi}^{T-1}) x_{T-1} + W_T(\bar{\xi}^{T-1}) x_T = h_T(\bar{\xi}^{T-1}), \\
& l_1 \le x_1 \le u_1, \\
& l_t(\bar{\xi}^{t-1}) \le x_t \le u_t(\bar{\xi}^{t-1}), \; t = 2, \ldots, T.
\end{aligned}
$$

Again, this value can be quite different from the here-and-know value and there can appear the same complication that is mentioned in remark 2.4.

Denote as $\bar{x}_1, \ldots, \bar{x}_T$ the optimal solution of the expected value problem. Similarly as in the two-stage problem, we now keep them fixed and we compute the true expectation of the objective function value for these decisions. The resulting value is

$$EEV = \mathrm{E}_{\xi^{T-1}} z(\bar{x}_1, \ldots, \bar{x}_T, \xi_1, \ldots, \xi_{T-1}).$$

3.2.3. Two-stage relaxation

Sometimes, we are not able to solve the given problem as multistage and we use its *two-stage relaxation.* In principle, we relax the nonanticipativity constraints in the second and other stages and we obtain a two-stage problem for T time periods

$$\begin{aligned} TP = \min_{x_1} \mathrm{E}_{\xi^{T-1}} \min_{x_2,\dots,x_T} & z(x_1,\dots,x_T,\xi_1,\dots,\xi_{T-1}) \\ \text{s.t. } & T_1 x_1 = h_1, \\ & T_2(\xi_1)x_1 + W_2(\xi_1)x_2 = h_2(\xi_1) \ \text{ a.s.}, \\ & T_3(\xi^2)x_2 + W_3(\xi^2)x_3 = h_3(\xi^2) \ \text{ a.s.}, \\ & \dots \\ & T_T(\xi^{T-1})x_{T-1} + W_T(\xi^{T-1})x_T = h_T(\xi^{T-1}) \ \text{ a.s.}, \\ & l_1 \le x_1 \le u_1, \\ & l_t(\xi^{t-1}) \le x_t \le u_t(\xi^{t-1}) \ \text{ a.s. for } t = 2,\dots,T. \end{aligned}$$

3.2.4. Rolling horizon

We can use another approach to multistage problems, which is based on rolling of time horizon. The here-and-now problem for time periods $1,\dots,T$ is solved first, giving an optimal solution $x_1^{*1},\dots,x_T^{*1}$. We use the component x_1^{*1} only, wait until ξ_1 is observed to be ξ_1^s and then we solve the here-and-now problem for time periods $2,\dots,T$ with the first stage decision fixed as x_1^{*1} and $\xi_1 = \xi_1^s$, i.e. we solve the problem

$$\begin{aligned} \min_{x_2} \mathrm{E}_{\xi_2|\xi_1^s} \min_{x_3} \dots \mathrm{E}_{\xi_{T-1}|\xi_{T-2},\dots,\xi_2,\xi_1^s} \min_{x_T} & z\Big(x_1^{*1},x_2,\dots,x_T,\xi_1^s,(\xi_2|\xi_1{=}\xi_1^s),\dots,(\xi_{T-1}|\xi_1{=}\xi_1^s)\Big) \\ \text{s.t. } & T_2(\xi_1^s)x_1^{*1} + W_2(\xi_1^s)x_2 = h_2(\xi_1^s), \\ & T_3(\xi^2|\xi_1=\xi_1^s)x_2 + W_3(\xi^2|\xi_1=\xi_1^s)x_3 = h_3(\xi^2|\xi_1=\xi_1^s) \ \text{ a.s.}, \\ & \dots \\ & T_T(\xi^{T-1}|\xi_1=\xi_1^s)x_{T-1} + W_T(\xi^{T-1}|\xi_1=\xi_1^s)x_T = h_T(\xi^{T-1}|\xi_1=\xi_1^s) \ \text{ a.s.}, \\ & l_2(\xi_1^s) \le x_2 \le u_2(\xi_1^s), \\ & l_t(\xi^{t-1}|\xi_1=\xi_1^s) \le x_t \le u_t(\xi^{t-1}|\xi_1=\xi_1^s) \ \text{ a.s.}, \ t = 3,\dots,T. \end{aligned}$$

This problem leads to an optimal solution $x_2^{*2},\dots,x_T^{*2}$. We wait until the realization of ξ_2 is known to be ξ_2^s and then we solve a here-and-now problem for the time periods $3,\dots,T$ with an objective function

$$z = z(x_1^{*1},x_2^{*2},x_3,\dots,x_T,\xi_1^s,\xi_2^s,(\xi_3|\xi_1=\xi_1^s,\xi_2=\xi_2^s),\dots,(\xi_{T-1}|\xi_1=\xi_1^s,\xi_2=\xi_2^s)).$$

This process is repeated with still shorter and shorter time horizon. The last problem which is solved is a deterministic (and static) problem

$$\begin{aligned} \min_{x_T} & z\Big(x_1^{*1},x_2^{*2},\dots,x_{T-1}^{*T-1},x_T,\xi_1^s,\dots,\xi_{T-1}^s\Big) \\ \text{s.t. } & T_T(\xi^{s,T-1})x_{T-1}^{*T-1} + W_T(\xi^{s,T-1})x_T = h_T(\xi^{s,T-1}), \\ & l_T(\xi^{s,T-1}) \le x_T \le u_T(\xi^{s,T-1}), \end{aligned}$$

where we denote $\xi^{s,T-1} = (\xi_1^s,\dots,\xi_{T-1}^s)$. Then we define

$$RH = \mathrm{E}_{\xi^{s,T-1}} z(x_1^{*1},x_2^{*2},\dots,x_T^{*T},\xi_1^s,\dots,\xi_{T-1}^s)$$

where $\mathrm{E}_{\xi^{s,T-1}}$ is expectation taken over all possible realizations of ξ^{T-1} and the resulting objective function value is denoted as RH (for *rolling horizon*). Further on, we will not write the random variables in the second and next steps explicitly conditionally on the previous realizations, but we always keep it in mind when dealing with the rolling horizon approach.

3.2.5. Dynamic formulation

Define *state vectors* $z_t \in Z_t \subseteq \mathbb{R}^{m_t}$, for $t = 1, \dots, T$, where z_t is a state of the system at the beginning of the time period t. In other words, z_t is the resulting state after applying the preceding decisions made in periods up to and including the $(t-1)$-th one and after observing realizations of $\xi_1, \dots, \xi_{t-1}$. The change from the state z_t to the state z_{t+1} under decision x_t and realization ξ_t is described by so called *transition function* $F_t \colon Z_t \times X_t \times \Xi_t \to Z_{t+1} \subseteq \mathbb{R}^{m_{t+1}}$ in the sense that $z_{t+1} = F_t(z_t, x_t, \xi_t)$. The initial state z_1 is given.

Denote the expected costs in the t-th period as $f_t = f_t(z_t, x_t) = \mathrm{E}_{\xi_t} \tilde{f}_t(z_t, x_t, \xi_t)$, where $\tilde{f}_t(z_t, x_t, \tilde{\xi}_t)$ expresses the costs in the t-th period under the realization $\tilde{\xi}_t$ of the random variable ξ_t, for all $\tilde{\xi}_t$ in the support Ξ_t. The total costs (or their worth after applying some negative-utility function) during the whole T-period process are expressed by an aggregate function $\Phi = \Phi[f_1, \dots, f_T]$. We will consider the most common shape of Φ which is $\Phi[f_1, \dots, f_T] = \sum_{t=1}^{T} f_t$ which is a sum of costs in all periods.

The whole *dynamic problem* then reads

$$
\begin{aligned}
DP = \min_{x_1,\dots,x_T} \ & \Phi[f_1(z_1, x_1), \dots, f_T(z_T, x_T)] \\
\text{s.t. } & z_1 \in Z_1, \ z_1 \text{ is given}, \\
& z_{t+1} = F_t(z_t, x_t, \xi_t), \ t = 1, \dots, T-1, \\
& x_t \in X_t, \ t = 1, \dots, T.
\end{aligned}
$$

Problems formulated in this way can be solved recurrently "backwards" as T nested problems with a parameter z denoting a state of the system.

This solution method uses Bellman's principle: If $(x_1^*, \dots, x_T^*)$ is a sequence of optimal decisions in the T-stage problem with initial state z_1, transition functions $F_1, \dots, F_T$ and costs functions $f_1, \dots, f_T$ then for any realization $\tilde{\xi}_1$ of ξ_1 is $(x_2^*, \dots, x_T^*)$ a sequence of optimal decisions in the $(T-1)$-stage problem with initial state $z_2 = z_2(\tilde{\xi}_1) = F_1(z_1, x_1^*, \tilde{\xi}_1)$, transition functions $F_2, \dots, F_T$ and costs functions $f_2, \dots, f_T$.

Denote

$$
\varphi_1(z_1; f_1) = \min_{x_1 \in X_1} f_1(z_1, x_1),
$$

and generally for $t = 2, \dots, T$ denote

$$
\begin{aligned}
\varphi_t(z_1; f_1, \dots, f_t; F_1, \dots, F_{t-1}) = \min_{x_1,\dots,x_t} \ & \sum_{\tau=1}^{t} f_\tau(z_\tau, x_\tau) \\
\text{s.t. } & x_\tau \in X_\tau, \ \tau = 1, \dots, t, \\
& z_{\tau+1} = z_{\tau+1}(\xi_\tau) = F_\tau(z_\tau, x_\tau), \ \tau = 1, \dots, t-1.
\end{aligned}
$$

According to the Bellman's principle, we have that

$$
\begin{aligned}
& \varphi_t(z_1; f_1, \dots, f_t; F_1, \dots, F_{t-1}) = \\
& \qquad = \min_{x_1 \in X_1} \left\{ f_1(z_1, x_1) + \varphi_{t-1}(F_1(z_1, x_1, \xi_1); f_2, \dots, f_t; F_2, \dots, F_{t-1}) \right\}.
\end{aligned}
$$

The new shape of the given problem is then

$$
\begin{aligned}
DP = \varphi_T\ & (z_1; f_1, \ldots, f_T; F_1, \ldots, F_{T-1}) \\
\text{s.t. } & x_t \in X_t,\ t = 1, \ldots, T, \\
& z_1 \text{ given}, \\
& z_{t+1} = z_{t+1}(\xi_t) = F_t(z_t, x_t, \xi_t),\ t = 1, \ldots, T-1.
\end{aligned}
$$

The problems for φ_t are problems with parameter z, which is the initial state of the system for φ_t. This state z depends on the preceding development of the random variables. In the backward recursion, at first the problem for $\varphi_1(z; f_T)$ is solved as a problem with parameter z, so the decision $x_T^*(z)$ is found. This decision minimizes the costs in the T-th period when the state of the system at the beginning of this period is z. Then the problem for $\varphi_2(z; f_{T-1}, f_T; F_{T-1})$ is solved, decision $x_{T-1}^*(z)$ is found as minimizing the expected costs in the $(T-1)$-th plus T-th period, when the state of the system at the beginning of the $(T-1)$-th period is z and the costs in the T-th period are minimized (by using the decision $x_T^*(z)$ for an appropriate state z). This solution process continues until the decision in the first period $x_1^*(z_1)$ (for the given state z_1) is reached. Then $x_1^*(z_1)$ is applied, ξ_1 observed as $\tilde{\xi}_1$, so that $z_2 = F_1(z_1, x_1^*(z_1), \tilde{\xi}_1)$ becomes the new state of the system. Then $x_2^*(z)$ is evaluated at z_2, $x_2^*(z_2)$ is applied, ξ_2 observed, the system transfers into new state, etc.

3.3. Value of information in multistage programming

In the preceding parts, we have introduced some approaches to solving a multistage problem: static formulation, two-stage relaxation, rolling horizon and dynamic formulation. We have reformulated the wait-and-see problem and the expected value problem introduced for two-stage problems. These various approaches result in various optimal objective function values. The goal of this paragraph is to show inequalities between these values similar to those presented for the two-stage problems, and consequently to define some types of value of information.

In [9], we can find a special dynamic formulation of some of the multistage problems described above (RP, WS and EEV), and basic inequalities (same as the ones presented in the next theorem) are also introduced there under some specific conditions. Some other approaches and values of information can be formulated under various assumptions, e.g. as in [12].

Theorem 3.1 For a T-stage stochastic problem (3.1), there is

$$WS \leq RP \leq EEV.$$

Proof. The proof would be the same as that for a two-stage problem, based on definition of feasible and optimal solutions. □

Remark 3.2 The inequality $EV \leq WS$ also holds here when the only random elements are right hand sides $h_2, \ldots, h_T$. The proof would be the same as that for a two-stage problem.

Theorem 3.3 For a T-stage stochastic problem (3.1), there is

$$TP \leq RP.$$

Proof. We can write

$$
\begin{aligned}
RP &= \min_{x_1} \mathrm{E}_{\xi_1} \left[\min_{x_2} \mathrm{E}_{\xi_2|\xi_1} \min_{x_3} \dots \mathrm{E}_{\xi_{T-1}|\xi^{T-2}} \min_{x_T} z(x_1, \dots, x_T, \xi_1, \dots, \xi_{T-1}) \right] \\
&= \min_{x_1} \mathrm{E}_{\xi_1} [r(x_1, \xi_1)]
\end{aligned}
$$

subject to constraints given in (3.1), and

$$
\begin{aligned}
TP &= \min_{x_1} \mathrm{E}_{\xi_1} \left[\mathrm{E}_{\xi_2,\dots,\xi_{T-1}} \min_{x_2,\dots x_T} z(x_1, \dots, x_T, \xi_1, \dots, \xi_{T-1}) \right] \\
&= \min_{x_1} \mathrm{E}_{\xi_1} [w(x_1, \xi_1)]
\end{aligned}
$$

subject to the same constraints, where we have defined the functions $r = r(x_1, \xi_1)$ and $w = w(x_1, \xi_1)$ as obvious. It is clear that for all fixed (x_1, ξ_1) (where x_1 is feasible and $\xi_1 \in \Xi_1$ is a possible realization) is $r(x_1, \xi_1)$ an optimal objective function value of a $(T-1)$-stage here-and-now problem for stages $2, \dots, T$. Similarly, $w(x_1, \xi_1)$ is an optimal objective function value of a wait-and-see problem for stages $2, \dots, T$. For any fixed (x_1, ξ_1), both of these $(T-1)$-stage problems are solved under the same constraints and it is clear that $w(x_1, \xi_1) \le r(x_1, \xi_1)$, which implies that $TP \le RP$. □

We will show that some other inequalities hold true. They are mentioned in [18] as well, but for one particular example only and with quite vague proofs.

Theorem 3.4 For a T-stage stochastic problem (3.1), there is

$$RH \le RP.$$

Proof. In the first step of the rolling horizon (RH) approach, we solve exactly the same problem as in the static (RP) approach. Hence, the sequence of optimal decisions $x_1^{*1}, \dots, x_T^{*1}$ which we obtain in the first step of the rolling horizon approach is the same as the sequence of optimal decisions $x_1^*, \dots, x_T^*$ obtained when solving the RP problem. The decisions x_t^{*t}, $t = 2, \dots, T$ obtained in the next steps of the RH approach differ from the decisions $x_t^{*1} = x_t^*,\ t = 2, \dots, T$, only if they lead to a better (i.e., lower) optimal objective function value than the decisions $x_t^{*1} = x_t^*,\ t = 2, \dots, T$. Hence, it cannot hold $RH > RP$. □

Theorem 3.5 For a T-stage stochastic problem (3.1), there is

(a) $$WS \le DP,$$

(b) $$WS \le RH.$$

Proof. When solving the wait-and-see problem, we know all future realizations of the random variables in advance. Hence, for any possible sequence $\tilde{\xi}^{T-1} = (\tilde{\xi}_1, \dots, \tilde{\xi}_{T-1})$ of realizations of $\xi_1, \dots, \xi_{T-1}$, we find a sequence of optimal decisions $x_1^{WS}, \dots, x_T^{WS}$ depending on $\tilde{\xi}^{T-1}$ and we compute the wait-and-see value as the expectation of the resulting objective function values over all possible $\tilde{\xi}$s. For an arbitrary $\tilde{\xi}^{T-1}$, we have

$$z(x_1^{WS}, \dots, x_T^{WS}, \tilde{\xi}^{T-1}) = \min_{x_1,\dots,x_T} z(x_1, \dots, x_T, \tilde{\xi}^{T-1}) \le z(\tilde{x}_1, \dots, \tilde{x}_T, \tilde{\xi}^{T-1}) \qquad (3.2)$$

for any feasible decisions $\tilde{x}_1, \ldots, \tilde{x}_T$.
(a) Especially, the inequality (3.2) holds for the decisions obtained as the optimal solutions of the dynamic problem. Hence, $\mathrm{E}_{\xi_1,\ldots,\xi_{T-1}} z(x_1^{WS}, \ldots, x_T^{WS}, \xi^{T-1}) \leq DP$, that is, $WS \leq DP$.
(b) The inequality (3.2) also holds for $\tilde{x}_1 = x_1^{*1}, \ldots, \tilde{x}_T = x_T^{*T}$, i.e. for the optimal solutions gained by the rolling horizon approach. Hence, we have that $WS \leq RH$. □

Remark 3.6 For a very special case of problem, inequality $DP \leq RH$ holds as well (see e.g. [18]). For this to be proven, the problem solved by the dynamic approach must be exactly the same as the one solved by the method of rolling horizon, which is not a priori assured because of the significant difference in formulations.

The preceding theorems allow us to complete the chain of inequalities:

$$WS \leq RH \leq RP \leq EEV.$$

Thanks to these inequalities, we can compare the results of different models for solving the same multistage problem. From the differences between the results, we can derive the value of the various approaches as compared with another ones.

Definition 3.7 Considering multistage problems and notation employed in this chapter, we define *value of stochastic solution* as $VSS = EEV - RP$,
value of rolling horizon as $VRH = EEV - RH$,
value of perfect information as $VPI = EEV - WS$.
All these values are nonnegative and they give the comparison of the true expectation of the objective function value when the optimal solutions of the expected value problem are used, with some of the other models benefiting from better and better usage of information, from the static stochastic model up to the clairvoyant's wait-and-see problem.

It is clear that the VPI creates an upper bound for the *expected value of perfect information* which is defined as

$$EVPI = RP - WS. \tag{3.3}$$

However, in the multistage problems it is quite debatable which one of DP, RH and RP should be compared to WS to define the expected value of perfect information. In our opinion this value, according to its name, should express the difference between expectation of the result of the clairvoyant's problem (which is WS) and expectation of the best possible result of the given problem formulated as nonanticipative, which is RH (or DP when the conditions required for $DP \leq RH$ are satisfied). Hence, we should define *expected value of perfect information for multistage problems* as $EVPI_M = RH - WS$ (or $EVPI_{\tilde{M}}$ as $EVPI_{\tilde{M}} = DP - WS$).
We can finally define *value of two-stage relaxation* as

$$VTR = RP - TP$$

which is also nonnegative thanks to the theorem 3.3.

3.4. Bounds on the value of information

In chapter 2.3., we have introduced GJEM bounds on value of information. We can generalize these bounds for multistage programs.

Similarly as in [11], we can also consider that a convex negative-utility function b applied on the expected optimal objective function value. Then we define a value of information $\tilde{V}$ in the same way as in the two-stage case.
Consider the situation when the decision maker takes a sequence of decisions $\{x_t\}$ in periods $t = 1, \dots, T$. After the t-th decision is taken, a random variable is observed to be ξ_t, so the process is nonanticipative and the order of making decisions and observing the random variables is $x_1, \xi_1, \dots, \xi_{T-1}, x_T$. We assume that $\xi_1, \dots, \xi_{T-1}$ are independent and for every t the support of ξ_t is Ξ_t.

The feasibility sets $K_1, \dots, K_T$ are now supposed to be fixed.
Net costs from decisions $\{x_t, t = 1, \dots, T\}$ and random realizations $\{\xi_t, t = 1, \dots, T-1\}$ is $z(x_1, \dots, x_T, \xi_1, \dots, \xi_{T-1}) = z(x^T, \xi^{T-1})$.

3.4.1. Linear negative-utility function

If the negative-utility function is linear the no-information problem is of the form

$$RP = Z_n^T = \min_{x_1 \in K_1} \mathrm{E}_{\xi_1} \min_{x_2 \in K_2} \dots \mathrm{E}_{\xi_{T-1}|\xi^{T-2}} \min_{x_T \in K_T} z(x^T, \xi^{T-1}) \tag{3.4}$$

and the perfect-information problem reads

$$WS = Z_p^T = \mathrm{E}_{\xi_1} \dots \mathrm{E}_{\xi_{T-1}} \min_{x_1 \in K_1} \dots \min_{x_T \in K_T} z(x^T, \xi^{T-1}).$$

In Z_n^T and Z_p^T, the upper index T stands for the number of stages, the subscripts n or p for no information or perfect information, respectively.
Now we can define the expected value of perfect information as

$$EVPI = Z_n^T - Z_p^T, \tag{3.5}$$

which is surely equivalent with the definition in (3.3). We now derive some bounds of the GJEM type. For simplicity, the next theorem is formulated for the case $T = 2$; generalization for arbitrary finite T is possible. We also do not write explicitly the decision variable x_3 which is of no use here; we can imagine that $z(x_1, x_2, \xi_1, \xi_2) = \min_{x_3} \tilde{z}(x_1, x_2, x_3, \xi_1, \xi_2)$ for an original objective function $\tilde{z}$.

Theorem 3.8 Suppose that $z = z(x^2, \xi^2)$ is a continuous function convex in (x^2, ξ^2), where $\xi^2 \colon \Xi^2 \to \langle a_1, b_1 \rangle \times \langle a_2, b_2 \rangle$, F_1, F_2 are marginal distribution functions of ξ_1, ξ_2, $x^2 \in K^2 = K_1 \times K_2$, K^2 is a compact and convex set, and $l = (l_1, l_2) \in \mathbb{N} \times \mathbb{N}$. Suppose for $k = 1, 2$ that $\langle a_k, b_k \rangle$ is subdivided at arbitrary points $a_k = d_0^k < d_1^k < \dots d_{l_k}^k = b_k$. Define

$$\alpha_{i_k}^k := \int_{d_{i_k-1}^k}^{d_{i_k}^k} dF_k(\xi_k), \; \beta_{i_k}^k := \frac{1}{\alpha_{i_k}^k} \int_{d_{i_k-1}^k}^{d_{i_k}^k} \xi_k dF_k(\xi_k), \; i_k = 1, \dots, l_k,$$

$$\delta_{i_k}^k := \alpha_{i_k}^k \left[\frac{\beta_{i_k}^k - d_{i_k-1}^k}{d_{i_k}^k - d_{i_k-1}^k} \right] + \alpha_{i_k+1}^k \left[\frac{d_{i_k+1}^k - \beta_{i_k+1}^k}{d_{i_k+1}^k - d_{i_k}^k} \right], \; i_k = 0, \dots, l_k,$$

$$\alpha_0^k = \alpha_{l_k+1}^k = \beta_0^k = \beta_{l_k+1}^k = d_{-1}^k := 0, \; k = 1, 2,$$

$$L_{l_1}(z) = \sum_{i_1=1}^{l_1} \alpha_{i_1}^1 z(x_1, x_2, \beta_{i_1}^1, \xi_2), \quad L_{l_2}(z) = \sum_{i_2=1}^{l_2} \alpha_{i_2}^2 z(x_1, x_2, \xi_1, \beta_{i_2}^2),$$

$$U_{l_1}(z) = \sum_{i_1=0}^{l_1} \delta_{i_1}^1 z(x_1, x_2, d_{i_1}^1, \xi_2), \quad U_{l_2}(z) = \sum_{i_2=0}^{l_2} d_{i_2}^2 z(x_1, x_2, \xi_1, d_{i_2}^2).$$

Then

(a) $$L_{n,l} := \min_{x_1 \in K_1} L_{l_1}(\min_{x_2 \in K_2} L_{l_2}(z)) \leq Z_n^2 \leq \min_{x_1 \in K_1} U_{l_1}(\min_{x_2 \in K_2} U_{l_2}(z)) =: U_{n,l},$$

(b) $$L_{p,l} := L_{l_1}(L_{l_2}(\min_{x^2 \in K^2} z)) \leq Z_p^2 \leq U_{l_1}(U_{l_2}(\min_{x^2 \in K^2} z)) =: U_{p,l},$$

(c) $$\max\{0; L_{n,l} - U_{p,l}\} \leq EVPI \leq U_{n,l} - L_{p,l}.$$

(d) If $\tilde{l} \geq l$ (i.e. $\tilde{l}_i \geq l_i,\ i = 1, 2$) and the partition corresponding to $\tilde{l}$ is at least as fine as that corresponding to l then the bounds in (c) corresponding to $\tilde{l}$ are at least as sharp as those corresponding to l.
(e) If each subinterval becomes arbitrarily small as $l_1 \to \infty$ and $l_2 \to \infty$ (which is denoted as $l \to \infty$), then

$$\lim_{l \to \infty} (U_{n,l} - L_{p,l}) = EVPI = \lim_{l \to \infty} (L_{n,l} - U_{p,l}).$$

Proof. The proof is based on application of theorem 2.23. For details see [11]. □

Theorem 3.9 If z is a continuous function of $(x^T, \xi^{T-1}) \in K^T \times \Xi^{T-1}$, where $\Xi^{T-1} = \Xi_1 \times \ldots \times \Xi_{T-1}$ and $K^T = K_1 \times \ldots \times K_T,\ K_t \subset \mathbb{R}^{n_t}$ are convex sets for all $t = 1, \ldots, T$, and z is convex in ξ^{T-1} for all fixed $x^T \in K^T$, then

$$0 \leq EVPI \leq \mathrm{E}_{\xi^{T-1}} z(\bar{x}^T, \xi^{T-1}) - z(\bar{x}^T, \bar{\xi}^{T-1}), \tag{3.6}$$

where $\bar{x}^T$ is an optimal solution to $\min\limits_{x^T \in K^T} z(x^T, \bar{\xi}^{T-1})$ and $\bar{\xi}^{T-1} = \mathrm{E}\xi^{T-1}$.

Proof. All minima and expectations exist and so $EVPI \geq 0$. Firstly, there is

$$\mathrm{E}_{\xi^{T-1}} z(\bar{x}^T, \xi^{T-1}) = \mathrm{E}_{\xi_1} \mathrm{E}_{\xi_2|\xi_1} \ldots \mathrm{E}_{\xi_{T-1}|\xi^{T-2}} z(\bar{x}^T, \xi^{T-1}) \geq Z_n^T, \tag{3.7}$$

because $\bar{x}^T$ is a feasible solution of the problem (3.4). Secondly,

$$Z_p^T = \mathrm{E}_{\xi^{T-1}} \min_{x^T \in K^T} z(x^T, \xi^{T-1}) \geq \min_{x^T \in K^T} z(x^T, \bar{\xi}^{T-1}) \tag{3.8}$$

since $\min\limits_{x^T \in K^T} z(x^T, \xi^{T-1})$ is a convex function of ξ^{T-1} and we can use Jensen's inequality. Using (3.7) and (3.8) in the same time, we obtain the inequality (3.6). □

3.4.2. Convex negative-utility function

Let us suppose that the decision maker solves a T-stage problem with a convex cost function $z = z(x^T, \xi^{T-1})$ and a convex and strictly increasing negative-utility function b. Now the nonanticipative problem reads

$$\min_{x_1 \in K_1} \mathrm{E}_{\xi_1} \min_{x_2 \in K_2} \ldots \mathrm{E}_{\xi_{T-1}|\xi^{T-2}} \min_{x_T \in K_T} b[z(x^T, \xi^{T-1})]. \tag{3.9}$$

Then *value of perfect information* $\tilde{V}$ is defined as a (unique, since b is strictly increasing) solution of the equation

$$\mathrm{E}_{\xi^{T-1}} \min_{x^T \in K^T} b[z(x^T, \xi^{T-1}) + \tilde{V}] = \min_{x_1 \in K_1} \mathrm{E}_{\xi_1} \min_{x_2 \in K_2} \ldots \mathrm{E}_{\xi_{T-1}|\xi^{T-2}} \min_{x_T \in K_T} b[z(x^T, \xi^{T-1})]. \tag{3.10}$$

Analogically as for two-stage problems, the left hand side of (3.10) represents the wait-and-see problem where $\tilde{V}$ is added to overall costs. This $\tilde{V}$ is the value of knowing the future in advance, or price of a research made to gain this knowledge, or price for postponing taking the decisions until realizations of all random variables are observed.

We can generalize the theorems formulated for the two-stage case and derive bounds on the value of information.

Theorem 3.10 Suppose that b is strictly increasing convex function on $\mathbb{R}$, $z = z(x^T, \xi^{T-1})$ is convex in (x^T, ξ^{T-1}) on the convex set $K^T \times \Xi^{T-1}$ where $K^T = K_1 \times \ldots \times K_T$, $K_t \subseteq \mathbb{R}^{n_t}$, ($x_t$ is a n_t-dimensional vector) for every t, and $\Xi^{T-1} \subseteq \mathbb{R}^{T-1}$, ($\xi_t$ is a scalar random variable for all t).
Then

$$0 \leq \tilde{V} \leq b^{-1}\left[\mathrm{E}_{\xi^{T-1}} b[z(\hat{x}^T, \xi^{T-1})]\right] - z(\hat{x}^T, \bar{\xi}^{T-1}),$$

where $\hat{x}^T$ is an optimal solution to $\min\limits_{x^T \in K^T} z(x^T, \bar{\xi}^{T-1})$ and $\bar{\xi}^{T-1} = \mathrm{E}\xi^{T-1}$.

Proof. The proof is analogical the proof of the theorem 2.26. □

Theorem 3.11 Suppose that
(1) b is strictly increasing and convex on $\mathbb{R}$ and z is convex jointly in all its arguments,
(2) $x^{*T} = x^{*T}(\xi^{T-1})$ is an optimal solution of $\min\limits_{x^T \in K^T} z(x^T, \xi^{T-1})$ and
$\hat{x}^T = (\hat{x}_1, \hat{x}_2(\xi_1), \ldots, \hat{x}_T(\xi^{T-1}))$ solves $\min\limits_{x_1 \in K_1} \mathrm{E}_{\xi_1} \ldots \mathrm{E}_{\xi_{T-1}|\xi^{T-2}} \min\limits_{x_T \in K_T} b[z(x^T, \xi^{T-1})]$,
$\Xi^{T-1} \subseteq \mathbb{R}^{S_1} \times \ldots \times \mathbb{R}^{S_{T-1}}$ (i.e. ξ_t is an S_t-dimensional real random vector),
(3) $z(x^{*T}, \xi^{T-1})$ and $z(\hat{x}^T, \xi^{T-1})$ have distributions belonging to the same family with two parameters that are independent functions of mean and variance, i.e. if $z(x^{*T}, \xi^{T-1}) \sim G(y; a_1, b_1)$ and $z(\hat{x}^T, \xi^{T-1}) \sim H(z; a_2, b_2)$, then $G(y) = H(z)$ whenever $\frac{y-a_1}{\sqrt{b_1}} = \frac{z-a_2}{\sqrt{b_2}}$, where $a_1,\ a_2$ are finite and $b_1,\ b_2$ are finite and positive.
Then

(a) $$\tilde{V} \geq \mathrm{E}_{\xi^{T-1}}\left[z(\hat{x}^T, \xi^{T-1}) - z(x^{*T}, \xi^{T-1})\right] \Leftrightarrow \operatorname{var} z(x^{*T}, \xi^{T-1}) \geq \operatorname{var} z(\hat{x}^T, \xi^{T-1}),$$

(b) $$\operatorname{var} z(x^{*T}, \xi^{T-1}) \geq \operatorname{var} z(\hat{x}^T, \xi^{T-1}) \Rightarrow \tilde{V} \geq EVPI,$$

where $\tilde{V}$ is the value of perfect information for the case of convex negative-utility function b (defined by (3.10)) and $EVPI$ is the expected value of perfect information for the case of linear or no negative-utility function (defined by (3.5)).

Proof. The proof is exactly same as the proof of the theorem 2.28. □

Remark 3.12 We can also define a value $\tilde{V}_M$ similar to $EVPI_M$ as a solution of the equation

$$\mathrm{E}_{\xi^{T-1}} \min_{x^T \in K^T} b[z(x^T, \xi^{T-1}) + \tilde{V}_M] = z^*_{RH} \tag{3.11}$$

where z^*_{RH} is the optimal objective function value gained when solving the same problem by the method of the rolling horizon. If the badness function b is convex and increasing then $\tilde{V}_M$ is nonnegative.

When the conditions required for the inequality $DP \leq RH$ to hold are satisfied, we can similarly define $\tilde{V}_{\tilde{M}}$ as a solution of the equation

$$\mathrm{E}_{\xi^{T-1}} \min_{x^T \in K^T} b[z(x^T, \xi^{T-1}) + \tilde{V}_{\tilde{M}}] = z^*_{DP} \tag{3.12}$$

where z^*_{DP} is the optimal objective function value gained when solving the problem in its dynamic formulation.

3.5. Value of information as a risk measure

In this part we will return to the approach used in chapter 2.9., generalized for a multistage case. We will again follow the ideas introduced in [16] and [15] to define risk processes and to show some properties of them.

Consider a random process $\xi_1, \dots, \xi_{T-1}$ defined on $(\Omega, \mathscr{F}, P)$ adopted to the filtration $\{\mathscr{F}_t, t = 1, \dots, T-1\}$ (which means that ξ_t is $\mathscr{F}_t$-measurable for $t = 1, \dots, T-1$), and $\mathscr{F}_0 = (\emptyset, \Omega)$.

Definition 3.13 A stochastic process $\{\xi_t, t = 1, \dots, T-1\}$ adapted to $\{\mathscr{F}_t, t = 1, \dots, T-1\}$ is called a *tree process*, if for $t = 1, \dots, T-1$ is $\mathscr{F}_t$ the σ-field generated by ξ_t.
As is proven in [15], a stochastic process is a tree process if and only if for all pairs $s < t$, $(\xi_s|\xi_t)$ which is the conditional distribution of ξ_s given ξ_t, is a constant almost surely (i.e. if we know the realization of ξ at the time t, we know all realizations of ξ up to time t).

Let $z = z(x_1, \xi_1, \dots, \xi_{T-1}, x_T)$ be a measurable real-valued objective (costs) function of $x_t,\ t = 1, \dots, T$ and $\xi_t,\ t = 1, \dots, T-1$. The x_ts are decision variables and the ξ_ts are uncertain costs in time periods $1, \dots, T$. We can again consider ξ_0 for technical reasons as a known constant almost surely. The distribution of all random variables is supposed to be known, and the decision process is nonanticipative. We now write the arguments of the function z in the order which reflects nonanticipativity of the process.

The multistage nonanticipative optimization problem (a problem of a common human decision maker who does not know the future random outcomes) can be written as

$$\begin{aligned} V^{\xi_1,\dots,\xi_{T-1}} = \min_{x_1,\dots,x_T} & \mathrm{F}\left[z(x_1, \xi_1, \dots, \xi_{T-1}, x_T)\right] \\ \text{s.t. } & x_t \in \mathscr{M}(\mathscr{F}_{t-1}, \Omega, \mathbb{R}),\ t = 1, \dots, T \end{aligned} \tag{3.13}$$

where F is a probability functional monotonic w.r.t. the first order stochastic dominance. Remind that for any σ-field $\mathscr{F}$, $\mathscr{M}(\mathscr{F}, \Omega, \mathbb{R})$ is a space of all $\mathscr{F}$-measurable functions from Ω to $\mathbb{R}$. The nonanticipativity constraint then has the meaning that $x_t = x_t(\xi_{t-1})$ can depend on the past realizations and past decisions only.
As expected, the anticipative (clairvoyant's) problem reads

$$\begin{aligned} V_C^{\xi_1,\dots,\xi_{T-1}} = \min_{x_1,\dots,x_T} & \mathrm{F}\left[z(x_1, \xi_1, \dots, \xi_{T-1}, x_T)\right] \\ \text{s.t. } & x_t \in \mathscr{M}(\mathscr{F}_{T-1}, \Omega, \mathbb{R}),\ t = 1, \dots, T. \end{aligned}$$

The risk involved in not having the full information (i.e. in the random income and the filtration $\{\mathscr{F}_t,\ t = 1, \dots, T-1\}$) is defined as

$$R^{\xi_1,\dots,\xi_{T-1}} = V^{\xi_1,\dots,\xi_{T-1}} - V_C^{\xi_1,\dots,\xi_{T-1}}.$$

$R^{\xi_1,\dots,\xi_{T-1}}$ is nonnegative, since the solution under full information gives an optimal objective function value which is lower or equal to the optimal objective function value gained by the nonanticipative solution, because the feasibility sets $\mathscr{M}(\mathscr{F}_{T-1}, \Omega, \mathbb{R})$ in the anticipative problem are larger or equal to the feasibility sets $\mathscr{M}(\mathscr{F}_{t-1}, \Omega, \mathbb{R})$ in the nonanticipative one.

If the function z is nondecreasing in variables $\xi_1, \ldots, \xi_{T-1}$ then $R^{\xi_1,\ldots,\xi_{T-1}}$ is monotonic in $\xi_1, \ldots, \xi_{T-1}$. This means that for two outcome streams $\{\xi_1, \ldots, \xi_{T-1}\}$ and $\{\zeta_1, \ldots, \zeta_{T-1}\}$ such that $\xi_t \le \zeta_t$ a.s. for all $t = 1, \ldots, T-1$ it holds $R^{\xi_1,\ldots,\xi_{T-1}} \le R^{\zeta_1,\ldots,\zeta_{T-1}}$.

Consider the same tree process $\{\xi_t, t = 1, \ldots, T-1\}$. For $s < t$ we can view ξ_s as the projection of ξ_t to the time t, i.e, $\xi_s = \mathrm{pr}_s(\xi_t)$. For $s > t$ we denote conditional distribution of ξ_s given ξ_t as $\xi_s|\xi_t$. The future process $\{\xi_s|\xi_t,\ s = t+1, \ldots, T-1\}$ is a tree process as well. Full conditional process given ξ_t is $\mathrm{pr}_1(\xi_t),\ \mathrm{pr}_2(\xi_t),\ \ldots,\ \mathrm{pr}_{t-1}(\xi_t),\ \xi_t,\ \xi_{t+1}|\xi_t,\ \ldots, \xi_{T-1}|\xi_t$. Denote $x_1^*, x_2^* = x_2^*(\xi_1), \ldots, x_T^* = x_T^*(\xi_{T-1})$ the (assume that unique) optimal solution of the nonanticipative problem (3.13). Now we can define so called *costs process:* Suppose that the value of ξ_t at time t is known and it is $\tilde{\xi}_t$, the decisions $x_1^*, \ldots, x_t^*$ up to time t were made in an optimal way; then the nonanticipative subproblem conditional at time t is

$$U_t = U_t(\tilde{\xi}_t) = \min_{x_{t+1},\ldots,x_T} \mathrm{F}\left[z(x_1^*, \mathrm{pr}_1(\tilde{\xi}_t), \ldots, \mathrm{pr}_{t-1}(\tilde{\xi}_t), x_t^*, \tilde{\xi}_t, x_{t+1}, \xi_{t+1}|\tilde{\xi}_t, \ldots, x_{T-1}, \xi_{T-1}|\tilde{\xi}_t)\right]$$
$$\text{s.t.} \quad x_{t+1} = x_{t+1}(\tilde{\xi}_t), \ldots, x_T = x_T(\xi_{T-1}|\tilde{\xi}_t).$$

(We can interpret this that a non-clairvoyant human being makes decisions at the time point t, when he already knows the realization ξ_t.) Now $\{U_t,\ t = 1, \ldots, T-1\}$ create a costs process.

Clairvoyant's anticipative problem in the same situation reads

$$U_{Ct} = U_{Ct}(\tilde{\xi}_t) = \min_{x_{t+1},\ldots,x_T} \mathrm{F}\left[z(x_1^*, \mathrm{pr}_1(\tilde{\xi}_t), \ldots, \mathrm{pr}_{t-1}(\tilde{\xi}_t), x_t^*, \tilde{\xi}_t, x_{t+1}, \xi_{t+1}|\tilde{\xi}_t, \ldots, x_T, \xi_{T-1}|\tilde{\xi}_t)\right]$$
$$\text{s.t.} \quad x_{t+1} = x_{t+1}(\xi_{T-1}|\tilde{\xi}_t), \ldots, x_T = x_T(\xi_{T-1}|\tilde{\xi}_t).$$

(Before realization of ξ_t a common human decision maker solves the problem and then he turns to be a clairvoyant.) Again, we consider $\{U_{Ct},\ t = 1, \ldots, T-1\}$ as a process. In each time period, feasibility set of the clairvoyant is a subset of the feasibility set of the common human decision maker and so $U_{Ct} \le U_t$ for all t; so it makes sense to define (*VPI*)*-process* as a process of value of perfect information (from the time t up to the end of the horizon) as $\{R_t,\ t = 1, \ldots, T-1\}$, where for $t = 1, \ldots, T-1$ is

$$R_t = U_t - U_{Ct} \ge 0.$$

We thus obtained a new and interesting information value type, or to be precise, a process of information values. Although it can be defined only for the multistage problems, we did not go too far from the results presented for the two-stage problems.

4. Conclusions

4.1. Information value types

We could see many times in this work that, in general, solving a stochastic problem in a situation when an information about future development of included random variables is available results in "better" optimal objective function value than solving the same problem without this information. It is also obvious that, even when the same information is available, we can choose among several approaches to the same problem and using different approaches, the (same) information is utilized to distinct extent. Different approaches thus result in generally distinct optimal solutions and in distinct optimal values of the objective function. Some inequalities were proven among the optimal objective function values. Differences (in the mathematical sense) between these values can be interpreted as a value of having a certain piece of information in the moment of solving the problem, or they express how much better is one method of solving the problem (or approaching to it) than another method (or approach) with the same level of available information.
It seems to be quite logical that "the best" approaches resulting to the lowest (for minimization problems) optimal objective function value are usually of the highest computational demands.
Especially for the multiperiod stochastic problems, we have shown many ways of approaching to the available information and of utilizing the increase of information in time. Again, knowledge of the future development of random variables included in the problem enables us to gain better results. Some differences between optimal objective function values resulting from different approaches to the problem are also obvious.

4.1.1. Classification of different types of value of information

In our opinion, the above mentioned types of information can be divided into two classes and one of them can be further divided into two subclasses, as is shown in the chart:

$$\text{All types of value of information} \begin{cases} \mathbf{A} \\ \mathbf{B} & \begin{cases} \mathbf{B_1} \\ \mathbf{B_2} \end{cases} \end{cases}$$

Below we present characterisation of the three classes and a list of information values which we categorise into each of the classes. Although we will deal with quantities that were already defined in previous chapters, we add their short characterisations to remind their definitions and meanings.

Values in the group **A** can be identified as differences (in the mathematical sense) between distinct optimal objective function values gained when using different approaches to a given stochastic problem with the same level of available information in all of the approaches. To be exact, these values do not represent a value of information. Rather, they can be interpreted as values of "better approach" to the given problem and of better ability to take advantage of the available information. Class **A** contains the following values:

Value of stochastic solution (chapter 2.1.6.) is defined as $VSS = EEV - RP$. The RP is the optimal objective function value of recourse problem (the nonanticipative one), while EEV is the true expectation of the objective function value gained when the first stage decision $x = x(\bar{\xi})$ is fixed as an optimal solution of the expected value problem, i.e. the problem where the random elements are replaced by their expectations. High values of

VSS signalise that it is much better to solve the (stochastic) recourse problem than the (deterministic) expected value problem, although the former one is more computationally demanding.

A redefinition of *VSS* (chap. 2.4.) is given as $VSS^u = EVRS^u - RP$, where $EVRS^u = \mathrm{E}_\xi z(\bar{x}^u, \xi)$ with $\bar{x}^u$ an optimal solution of a related deterministic problem with fixed reference scenario ξ^u. For $\xi^u = \bar{\xi}$ it holds that $EVRS^u = EEV$ and $VSS^u = VSS$. The notion of VSS^u is useful for comparison to some other values related to pairs subproblems.

For multistage problems, we have defined some specific information values regarding the approach to the problem and utilisation of the available information (chap. 3.3.).
Value of rolling horizon $VRH = EEV - RH$ expresses how much we can save when using the rolling horizon approach instead of solving the (deterministic) expected value problem and using its solution in a real multistage stochastic problem.
Some other values expressing comparison of different approaches to the same multistage problem can be defined as $RP - RH$ or as $RH - DP$ which can only be defined under conditions which assure that exactly the same problem is formulated in the rolling horizon approach and in the dynamic approach.

Class **B** covers all types of values which are indeed values of information. Each of them is calculated as a difference between two optimal objective function values, from which the first one (the greater) is the resulting optimal objective function value gained without a piece of information and the second one (the lower) is obtained when computing the same problem when this piece of information is available. The optimal objective function value obtained without an information on future development of some random variables is a function of their concrete realization, and so it is necessary to add the expected value operator E or an other functional F. Hence, we speak about different types of expected value of information.
The class **B** can be further divided into subclasses $\mathbf{B_1}$ and $\mathbf{B_2}$.

The subclass $\mathbf{B_1}$ covers expected values of information that has to do with future development of random variables included in the given problem. They are represented by the following values:
Expected value of perfect information (chap. 2.1.6.) is defined as $EVPI = RP - WS$, where *RP* is the optimal objective function value of the recourse problem obtained when solving it as a nonanticipative problem, and *WS* is the expected optimal objective function value of the same problem solved with full information on the future development of random variables included in the problem. High values of *EVPI* signalise that it could be worth doing to find out as much as possible about the development of the random variables. *EVPI* is one of the most often referred characteristics.

We have defined *expected value of perfect information for chance-constrained problem* (chap. 2.8.1.) as $EVPI^C = RP - WS$, where *RP* and *WS* are defined by (2.48) and (2.49), respectively. The intuition about $EVPI^C$ is the same as the one about *EVPI*.

Having full information on the future development of all random variables leads to an anticipative wait-and-see formulation of the (two-stage or multistage) approach. We can then define *value of perfect information* (chap. 3.3.) as $VPI = EEV - WS$. In fact, this information value belongs to the class $\mathbf{B_1}$ as well as to the class **A** because it compares the result of the wait-and-see approach to the optimal objective function value gained by the simplest and least computationally demanding *EEV* approach.

Definition 2.16 gives a mathematical precision to an intuitive feeling of what is an information structure (chap. 2.3.1.). An information structure η provides a partial (or full) information under which a decision maker gains an optimal objective function value Z_η of the given two-stage program. This Z_η is lower or equal to the optimal objective function value $Z_n = RP$ gained without any such information. *Value of partial information (given by η)* is then defined as $V_\eta = Z_n - Z_\eta$. Its properties are quite natural, they are connected with features of conditional distributions and expectations.
Immediately, *value of increasing partial information (from η^1 to finer η^2)* is defined as $V_{\eta^1} - V_{\eta^2}$.

If the "badness" given by costs $z(x, \xi)$ is not equal to the costs, we use a negative-utility function b (chap. 2.3.3.). The negative-utility of costs $z(x, \xi)$ is then $b[z(x, \xi)]$ which is also minimized. Instead of $EVPI$, we define *value of information* $\tilde{V}$ as a solution of the equation $\mathrm{E}_\xi \min b[z(x, \xi) + \tilde{V}] = \min \mathrm{E}_\xi b[z(x, \xi)]$. High values of $\tilde{V}$ signalise that the value of full information on the realization on ξ is high, and vice versa.

$\tilde{V}_\eta$ is *value of partial information (given by an information structure η)* in a problem with a strictly increasing convex negative-utility function b (chap. 2.3.4.). The notion of $\tilde{V}_\eta$ connects logically the value of partial information V_η (defined for problems with a linear negative-utility function) and value of full information $\tilde{V}$ (defined for problems with convex negative-utility function b).

In the modified wait-and-see approach (chap. 2.6.), we set $MWS_i = \mathrm{E}_\xi z(x_i^*, \xi)$, where x_i^* is an optimal first stage solution of the related deterministic problem for one fixed scenario ξ^i. *Expected value of perfect information in the modified wait-and-see approach* is then defined as $MEVPI = \min_{i:\xi^i \in \Xi} MWS_i - WS$. It creates an upper bound on $EVPI$ and its value shows worth of having full information on the realization of ξ as compared to the situation when we do not have this information and we also do not solve the problem as a stochastic one.
$MEVPI_N$ ($N = 1, \dots, K$ where K is the number of possible scenarios) is a generalization of *MEVPI*. It is defined as $MEVPI_N = \min_{S \subseteq \{1,\dots,K\}, \#S=N} MWS_S - WS$, where $MWS_S = \mathrm{E}_\xi z(x_S^*, \xi)$, x_S^* is an optimal first stage decision for the subset S of the set of all possible scenarios. $MEVPI_N$ has a similar interpretation as *MEVPI*.

When representing the available information via σ-fields (chap. 2.9.), we can define *expected value of partial information given by a σ-field $\mathscr{F}$* as $V^\xi_{\mathscr{F}_0} - V^\xi_{\mathscr{F}}$, where $\mathscr{F}_0 = \{\emptyset, \Omega\}$ is the simplest σ-field. Now, $V^\xi_{\mathscr{F}_0}$ is an optimal objective function value of a problem solved under no information $\mathscr{F}_0$ (so that the optimal solution x is $\mathscr{F}_0$-measurable). $V^\xi_{\mathscr{F}}$ is an optimal objective function value of the same problem solved when a partial information represented by the σ-field $\mathscr{F}$ is available (so that the optimal solution x is $\mathscr{F}$-measurable).
An information value which is complementary to the last mentioned one is *risk value of the information represented by the σ-field $\mathscr{F}$*. It is defined as $R^\xi_{\mathscr{F}} = V^\xi_{\mathscr{F}} - V^\xi_{\tilde{\mathscr{F}}}$, where $\tilde{\mathscr{F}}$ is the largest σ-field representing the full information.

A generalization of expected and risk value of perfect information ($EVPI_\mu$, $RVPI_\varrho$) is outlined in a remark 2.62.

For multistage problems (chap. 3.3.), special *value of perfect information for multistage problems* is defined as $EVPI_M = RH - WS$. It compares optimal objective function values

resulting from the (anticipative) wait-and-see approach and from the "best" nonanticipative approach which we have formulated. If the conditions assuring that exactly the same problem is solved using the dynamic approach and the static approach (which leads to RP) then we also define its analogy $EVPI_{\tilde{M}}$ as $EVPI_{\tilde{M}} = DP - WS$. Of course, common *expected value of prefect information* is defined for the multistage problems as well.
Analogically as in the case of two-stage problems, expected value of perfect information $EVPI$ can be also defined as $EVPI = Z_n^T - Z_p^T$, where the superscripts T denotes the number of stages of the problem, and the subscripts n and p stand for no-information structure and perfect information structure, respectively (chap. 3.4.1.).

Value of two-stage relaxation (chap. 3.3.) which is defined as $VTR = RP - TP$ is a value of information which allows us to relax the nonanticipativity constraints in the second and other stages of the multistage problem in order to formulate the problem as a two-stage one.

When we generalize the multistage problem by usage of a convex negative-utility function b (chap. 3.4.2.), we can define *value of perfect information* $\tilde{V}$ as a solution of the equation (3.10), where the left side of the equation represents the WS approach with additional costs $\tilde{V}$, and the right side represents the RP approach.
In the same situation, an analogy to $EVPI_M$ is $\tilde{V}_M$ which is defined as a solution of the equation (3.11). The left side of the equation again represents the WS approach with additional costs $\tilde{V}_M$, and the right side represents the RH approach.

When we use the approach to multistage problems as in chapter 3.5., we can define *risk involved in not having full information* as $R^{\xi_1,\dots,\xi_{T-1}} = V^{\xi_1,\dots,\xi_{T-1}} - V_C^{\xi_1,\dots,\xi_{T-1}}$ where $V^{\xi_1,\dots,\xi_{T-1}}$ is an optimal objective function value of a nonanticipative problem $\min_{x_1,\dots,x_T} \mathrm{F}[z(x_1, \xi_1, \dots, \xi_{T-1}, x_T)]$, while $V_C^{\xi_1,\dots,\xi_{T-1}}$ is an optimal objective function value of the same problem with fully anticipative constraints. The idea of this information value is analogical to the idea of $EVPI$, but now E is replaced by a functional F and nonanticipativity is expressed via measurability w.r.t. σ-fields.
For the same problem, we have defined a *VPI-process* $\{R_t,\ t = 1, \dots, T-1\}$ which is a sequence of nonnegative elements $R_t = U_t - U_{Ct}$. Here, U_t and U_{Ct} is a nonanticipative and anticipative optimal objective function value, respectively, both conditionally to the realization of ξ_t in time t.

Let us focus on subclass $\mathbf{B_2}$ now. This subclass covers expected values of information about a distribution of some random variables. This kind of problem is not in a focus of the main interest in this work and it is modelled in the way that in fact it could also belong to $\mathbf{B_1}$. Distribution functions of some random variables depend on an unknown parameter which is again considered as a random variable. Having the information about a future development of this random parameter, we know exactly the distribution functions of the random variables in the "first level." The distribution of these variables is not known entirely without this information and, intuitively said, the extent of uncertainty having its source in the "first level" variables is multiplied by the uncertainty arising from randomness of the parameter. In scenario-based problems this increases the number of scenarios we have to hedge against. Then, in average, the optimal solution fits worse to all particular scenarios. It is therefore useful to know completely the distribution of all random variables (which is usually assumed). The values belonging to the subclass $\mathbf{B_2}$ are as follows:

$EVSI_n$ is an *expected value of sample information* of a sample of size n (chap. 2.7.). This information value is defined as a difference between two optimal objective function values: at first, given problem is solved without full knowledge on the distribution of the random variable ξ; secondly, the same problem is solved when a supportive sample of size n from the ξ's distribution is available. The sample gives a (partial) information on the true distribution of ξ. Larger samples give more valuable information. When the sample size n tends to infinity and the limiting sample provides perfect information on the distribution of ξ, then $EVSI_\infty$ is defined as *expected value of sample information* of the infinite sample.
For chance-constrained problems with the same complication (i.e., the distribution of ξ is not fully known and some information on it can be gained by sampling), $EVSI_n^C$ and $EVSI_\infty^C$ are defined analogically, as well as $EVSI_n^{JC}$ and $EVSI_\infty^{JC}$ for problems with joint chance-constraints (chap. 2.8.2.).

4.1.2. Common features of individual types of value of information

In this work, we have defined many different types of information values. Some of them can be seen simply as a generalization of some other ones, but another ones are totally different. In this section, we intend to focus on common features of all of the information values, or at least most of them.
The first (and for us, the most important) feature which we have proven for all the defined information values is nonnegativity. This is in accordance with an intuitive feeling that we are able to solve a given stochastic problem when (full or partial) information is available at least as well as without this information, no matter whether we speak about information on distribution of random variables included in the problem or information on future development of the random variables.
An intuitive interpretability is also very useful property. Usually, it is quite clear from definition what is the meaning of the defined information value. This is quite important when deciding what indicates high, low or zero value of the information value. Low values signalise that it is not worth doing to gain the information (for information value types from class **B**), or that it is not worth doing to use the more sophisticated approach to solving the problem (for information value types from class **A**).
The information value types which we have categorised into class **A** are defined as a difference of two optimal objective function values: the greater one is gained when using an approach which poorly utilises the available information, while the lower one results from usage of an approach which utilises the information more effectively. This "better" approach is usually more computationally demanding. To compute the information value, we would have to solve the problem with help of the both of the approaches, which is not reasonable. The information value types from class **B** are also defined as a difference of two optimal objective function values: the greater one belongs to a nonanticipative problem (the decision maker does not use any information which is not available in the time of making the decision), while the lower one is gained when solving the anticipative problem (we say that "the decision maker is a clairvoyant" as he uses some information which cannot be available in the time of making the decision). To compute this information value, we would have to solve both of the problems again. It is often quite computationally demanding.
For the reasons given in the preceding paragraph, it is quite useful that we can often found a lower or/and upper bounds on the information value, which do not require computation of both the values from the definition of the given information value.

The last property we would like to point out is that different information values are not generally comparable. Their meanings can be very different (e.g. *EVPI* and *VSS*) and hardly any inequalities work between information values from different classes. Therefore, high (or low) information value of one type usually does not imply high information value of another type.

In general, we can conclude that the resulting optimal objective function value depends on formulation of a given problem, which results in existence of information value types in the class **A**. It also depends on information that is available in the time of making the decision, which results in existence of information value types in the class **B**.
However, it is useful to keep in mind that our results were gained under certain conditions on the structure of problems—we usually considered linear problems only, sometimes with additional properties which were to various extent restrictive.

References

[1] Artzner P., Delbaen F., Eber J.-M., Heath D. (1999): Coherent measures of risk. *Mathematical Finance* **9**, 203–228.

[2] Avriel M., Williams A. C. (1970): The Value of Information and Stochastic Programming. *Operations Research* **18**, 947–954.

[3] Birge J. R. (1982): The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical Programming* **24**, 314–325.

[4] Birge J. R., Louveaux F. (1997): Introduction to Stochastic Programming. Springer-Verlag, New York.

[5] DeGroot M. H. (1970): Optimal Statistical Decisions. McGraw Hill, New York.

[6] Dupačová, J. (1986): Stochastické programování. MŠMT, Praha.

[7] Gunderson H. S., Morris J. G., Thompson H. E. (1978): Stochastic programming with recourse: A modification from an applications viewpoint. *Journal of the Operational Research Society* **29/8**, 769–778.

[8] Hanoch G., Levy H. (1969): The efficiency analysis of choices involving risk. *The Review of Economic Studies* **36**, 335–346.

[9] Hartley R. (1980): Inequalities for a Class of Sequential Stochastic Decision Processes. In: *Stochastic Programming.* M. A. H. Dempster (Ed.), Academic Press, London, 109–123.

[10] Hausch D. B., Ziemba W. T. (1983): Bounds on the value of information in uncertain decision problems II. *Stochastics* **10**, 181–217

[11] Huang C. C., Vertinsky I., Ziemba W. T. (1977): Sharp bounds on the value of perfect information. *Operations Research* **25**, 128–139.

[12] Huang K., Ahmed S. (2005): The value of multi-stage stochastic programming in capacity planning under uncertainty. *SPEPS 2005/15.*

[13] Jagannathan R. (1985): Use of sample information in stochastic recourse and chance-constrained programming models. *Management Science* **31**, 96–108.

[14] Lavalle I. H. (1986): On information-augmented chance-constrained programs. *Operation Research Letters* **5**, 225–230.

[15] Pflug G. Ch.(2004): The value of perfect information as a risk measure. In: *Dynamic Stochastic Optimization.* K. Marti, Y. Ermoliev, G. Pflug (Eds.), Springer-Verlag, Berlin, 2004, 275–292.

[16] Pflug G. Ch., Ruszczynski A. (2005): Measuring risk for income streams. *Computational Optimization and Applications* **32**, 161–178.

[17] Ruszczynski A., Shapiro A. (2004): Optimization of convex risk functions. *Optimization Online*, to appear in Mathematics of Operations Research.

[18] Wagner, J. M., Berman O. (1995): Models for planning capacity expansion of convenience stores under uncertain demand and the value of information. *Annals of Operations Research* **59**, 19–44.

[19] Wets R. J.-B. (2000): Stochastic programming models: Wait-and-see versus here-and-now. *Stochastic Programming Resources.*

Corrigendum

Correction of chapter 2.8.1. Value of information in chance-constrained programming of the diploma thesis "Očekávaná hodnota informace ve stochastickém programování."

Chapter 2.8.1. is devoted to formulation of here-and-now (RP) and wait-and-see (WS) chance-constrained problems and also to derivation of expected value of perfect information ($EVPI$) and proof of its nonnegatvity. As there are some mistakes (mainly due to unsuitable transcriptions), the purpose of this Annex is to correct these mistakes and to give formulation of the RH and WS problems and $EVPI$ in the chance-constrained programming in a proper way. For the sake of transparency, the problems will be written in a more detailed way and the penalization function used in the diploma thesis will be avoided.
In spite of the incorrect transcriptions, the idea used in the diploma thesis is correct and all the presented general conclusions are true. Also, Remark 2.55 still holds.

To explain our motivation, consider a simple problem

$$\begin{aligned} RP = \min_x \ & z(x) \\ \text{s.t. } & x \in K, \\ & x \geq \xi \ \text{ a.s.}, \end{aligned} \tag{A.1}$$

where z is a real convex function bounded on K, K is an interval in $\mathbb{R}$ and ξ is a real scalar random variable with known distribution and support Ξ. The second contraint in (A.1) gives us that $x \geq \xi$ a.s., that is $P(x \geq \xi) = 1$. If we partly relax this constraint, we obtain a *chance-constrained here-and-now problem*

$$\begin{aligned} RP = \min_x \ & z(x) \\ \text{s.t. } & x \in K, \\ & P(x \geq \xi) \geq \alpha \end{aligned} \tag{A.2}$$

for some $\alpha \in \langle 0; 1\rangle$ given. Now the constraint has the meaning that $x \geq \xi$ "sufficiently often" – at least with probability α. For $\alpha = 1$, problems (A.1) and (A.2) are equivalent. Consider the chance-constrained problem (A.2), now with ξ with discrete distribution with $P(\xi = \xi^s) = p^s$, $s = 1, \ldots, S$, $\sum_{s=1}^{S} p^s = 1$ and $p^s \geq 0 \ \forall s$. The problem can be equivalently written as

$$\begin{aligned} RP = \min_x \ & z(x) \\ \text{s.t. } & x \in K, \\ & \sum_{s: x \geq \xi^s} p^s \geq \alpha. \end{aligned} \tag{A.3}$$

Using the idea of explicite nonanticipativity constraints, we can rewrite (A.3) as

$$\begin{aligned} RP = \min_{\hat{x}, x^s: s=1,\ldots,S} \ & \sum_{s=1}^{S} p^s . z(x^s) \\ \text{s.t. } & x^s \in K, \ s = 1, \ldots, S, \\ & \sum_{s: x^s \geq \xi^s} p^s \geq \alpha, \\ & x^s = \hat{x}, \ s = 1, \ldots, S, \ \hat{x} \in K. \end{aligned} \tag{A.4}$$

As $x^s = \hat{x} \ \ \forall s$, no expectation is in fact needed there.
Related *wait-and-see problem* is then obtained by relaxation of the nonanticipativity constraint and by placing the expectation in front of the minimum:

$$\begin{aligned} WS = \sum_{s=1}^{S} p^s . \min_{x^s : s=1,\ldots,S} & \ z(x^s) \\ \text{s.t. } & x^s \in K, \ s = 1, \ldots, S, \\ & \sum_{s : x^s \geq \xi^s} p^s \geq \alpha. \end{aligned} \tag{A.5}$$

More generally, problems (A.4) and (A.5) can be formulated for ξ random variable with a distribution function F as

$$\begin{aligned} RP = \min_{\hat{x}, x_\xi : \xi \in \Xi} & \ z(x_\xi) \\ \text{s.t. } & x_\xi \in K \ \text{ a.s.}, \\ & \int_{\{\xi : x_\xi \geq \xi\}} dF(\xi) \geq \alpha, \\ & x_\xi = \hat{x} \ \text{ a.s.}, \ \hat{x} \in K. \end{aligned} \tag{A.6}$$

Now the decision x_ξ depends on ξ, but the nonanticipativity constraint makes it nonrandom again. The related wait-and-see problem is then

$$\begin{aligned} WS = \mathrm{E}_\xi \min_{x_\xi : \xi \in \Xi} & \ z(x_\xi) \\ \text{s.t. } & x_\xi \in K \ \text{ a.s.}, \\ & \int_{\{\xi : x_\xi \geq \xi\}} dF(\xi) \geq \alpha. \end{aligned} \tag{A.7}$$

Theorem A.1 Consider the chance-constraint problem (A.6) and its wait-and-see modification (A.7). Define *expected value of perfect information* for these problems as

$$EVPI^C = RP - WS.$$

Then $EVPI^C$ is nonnegative.
Proof. Let x^* be an optimal decision of problem (A.6). Constant function $x_\xi = x^*$ a.s. is a feasible solution of (A.7) and so the optimal solution x_ξ^* of (A.7) (for any particular realization of ξ) differs from x^* only if $z(x_\xi^*) \leq z(x^*)$. Hence, $WS \leq RP$. □

It is easy to find an example for $WS < RP$: consider $z(x) = x, \ K = \mathbb{R}, \ \alpha = 0,8$ and ξ with a uniform distribution on $\langle 0; 1 \rangle$. Then $RP = 0,8$. Function $x_\xi = \xi$ is a feasible solution of the related wait-and-see problem since $\int_{\{\xi : x_\xi \geq \xi\}} dF(\xi) = 1 \geq 0,8$ for this function x_ξ. Hence, $WS \leq \mathrm{E}\xi = 0,5 < 0,8 = RP$.

Note that a formulation with help of a penalisation function (similar to the one introduced in the diploma thesis) is not difficult and is usually very useful when solving numerically a particlar example, not only in the chance-constraint programming.

Prague, 24.11.2006