Springer Series in Operations Research and Financial Engineering

John R. Birge
François Louveaux

Introduction to Stochastic Programming

Second Edition

Springer

Springer Series in Operations Research and Financial Engineering

Series Editors:
Thomas V. Mikosch
University of Copenhagen
Laboratory of Actuarial Mathematics
DK-1017 Copenhagen
Denmark
mikosh@act.ku.dk

Sidney I. Resnick
Cornell University
School of Operations Research and
Industrial Engineering
Ithaca, NY 14853
U.S.A.
sirl@cornell.edu

Stephen M. Robinson
University of Wisconsin-Madison
Department of Industrial Engineering
Madison, WI 53706
U.S.A.
smrobins@wisc.edu

For further volumes:
http://www.springer.com/series/3182

John R. Birge • François Louveaux

Introduction to Stochastic Programming

Second Edition

Springer

John R. Birge
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, Illinois 60637
USA
john.birge@chicagobooth.edu

François Louveaux
Department of Business Administration
University of Namur
Rempart de la Vierge 8
B-5000, Namur
Belgium
francois.louveaux@fundp.ac.be

ISSN 1431-8598
ISBN 978-1-4614-0236-7 e-ISBN 978-1-4614-0237-4
DOI 10.1007/978-1-4614-0237-4
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011929942

Mathematics Subject Classification (2010): 37N40, 46N10, 49L20, 49Mxx (all), 49N30, 49N15, 90-01, 90B50, 90C05, 90C06, 90C15, 90C39

© Springer Science+Business Media, LLC 2011
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Richard and Joelle,
Sebastien, Jérôme, Quentin, and Géraldine.

Preface

Since the publication of the first edition of this book, we have been encouraged by the growing interest in stochastic programming and its application in a variety of areas, including routine use in many industries from transportation and logistics to finance and energy. We have also been heartened by the many new methodological and theoretical advances within the field. In this second edition, we have attempted to capture aspects of both recent applications and models as well as new practically relevant methods and theory. As in the first edition, our primary goal is to provide students and other readers with an appreciation of how to build uncertainty into an optimization model, what differences in decisions might result from recognizing the presence of uncertainty, and how and what kinds of models are amenable to solution. We have focused the second edition on satisfying these main objectives while also uncovering basic research questions to give beginning researchers a foundation upon which to build more in-depth knowledge.

To help make the relevant issues in modeling, solving, and analyzing stochastic programs more evident, we have incorporated more examples than in the first edition so that the each of the main modeling, solution, and analysis processes are illustrated with a detailed example. We have also added many exercises whose solutions provide additional insights into stochastic programming concepts and tools. Many of these exercises assume the availability of software to solve basic linear and nonlinear optimization models and to construct algorithmic procedures involving matrix operations. Since we view completing these exercises as a key part of understanding the material, instructors should ensure that students have adequate programming skills to implement the methods described in the book.

Besides additional examples and exercises throughout the book, we have reorganized the material to improve the logical flow and to eliminate unnecessary or complicating issues before explaining the most practically relevant material. Specific changes in the second edition include the following:

- a new section (Section 1.5) and routing example in Chapter 1;
- a worked-out modeling exercise (Section 2.8) and a section on risk modeling and robust formulation (Section 2.9 in Chapter 2;

- re-arrangement and simplification of the material in Chapter 3 to emphasize basic model characteristics and illustrate them with examples;
- complete re-organization and combination of Chapters 5 and 6 into a new Chapter 5 that unifies the treatment of cutting-plane methods and again provides additional examples;
- an additional section on Lagrangian multistage methods in Chapter 6 (formerly Chapter 7);
- a completely re-organized version of Chapter 7 (formerly Chapter 8) including new methods and review material on combinatorial optimization;
- additional examples in Chapter 8 (formerly Chapter 9) including bounds on loss probabilities in loan portfolios;
- re-organization of Chapter 9 (formerly Chapter 10) to place practical methods earlier and to include a new section on Monte Carlo methods for probabilistic constraints;
- re-organization of Chapter 10 (formerly Chapter 11) to include new sections on scenario generation, multistage sampling methods, and approximate dynamic programming methods;
- removal of the short chapter (formerly Chapter 12) on a capacity expansion case study.

We anticipate that classes would follow much of the same sequence as we suggested for the first edition, but, with the increased availability of software to implement methods, we recommend that instructors include more computational exercises and additional modeling projects to fit students' interests. Any course should again start with the first two chapters to provide the application and modeling context. Depending on student interest, a typical class would generally include Chapters 3, 4, and Sections 5.1, 5.2, and 5.5 to present the most typical types of methods. For basic approximations, a modeling-focused class could focus on the main techniques in Chapters 8, 9, and 10 (for dynamic models), while a theoretically-oriented class might emphasize the analytical results in those chapters. A more computationally focussed class might emphasize the remainder of Chapter 5 plus Chapters 6 and 7.

We wish to thank the many people who sent us comments and suggestions about the first edition of the book and the numerous students we have worked with and all those who have helped us see stochastic programming from a fresh perspective every time we encounter something new. Among the many who have contributed, we thank Michael Dempster, Michel Gendreau, Maarten van der Vlerk, and Bill Ziemba. Thanks are also due to Martine Van Caeneghem for her patient typing of the modifications in Namur. We also again thank Fonds National de la Recherche Scientifique, the National Science Foundation, as well as the U.S. Department of Energy, and the University of Chicago Booth School of Business for their financial support.

In our first edition, we finished the preface with special thanks to our wives, Pierrette and Marie, to whom our book was dedicated. These thanks are more than ever very much present in our hearts. Now, we also want to express our proudness and joy of having such great children. We have thus decided to dedicate this second edition to them. We may thus expect that the third edition will be dedicated to our grandchildren, although the timing of this edition and the number of lines needed for this future dedication remain unknown.

Chicago, Illinois, USA John R. Birge
Namur, Belgium François Louveaux

Preface to the First Edition

According to a French saying “Gérer, c’est prévoir,” which we may translate as “(The art of) Managing is (in) foreseeing.” Now, probability and statistics have long since taught us that the future cannot be perfectly forecast but instead should be considered random or uncertain. The aim of stochastic programming is precisely to find an optimal decision in problems involving uncertain data. In this terminology, *stochastic* is opposed to *deterministic* and means that some data are random, whereas programming refers to the fact that various parts of the problem can be modeled as linear or nonlinear mathematical programs. The field, also known as *optimization under uncertainty*, is developing rapidly with contributions from many disciplines such as operations research, economics, mathematics, probability, and statistics. The objective of this book is to provide a wide overview of stochastic programming, without requiring more than a basic background in these various disciplines.

Introduction to Stochastic Programming is intended as a first course for beginning graduate students or advanced undergraduate students in such fields as operations research, industrial engineering, business administration (in particular, finance or management science), and mathematics. Students should have some basic knowledge of linear programming, elementary analysis, and probability as given, for example, in an introductory book on operations research or management science or in a combination of an introduction to linear programming (optimization) and an introduction to probability theory.

Instructors may need to add some material on convex analysis depending on the choice of sections covered. We chose not to include such introductory material because students’ backgrounds may vary widely and other texts include these concepts in detail. We did, however, include an introduction to random variables while modeling stochastic programs in Section 2.1 and short reviews of linear programming, duality, and nonlinear programming at the end of Chapter 2. This material is given as an indication of the prerequisites in the book to help instructors provide any missing background. In the Subject Index, the first reference to a concept is where it is defined or, for concepts specific to a single section, where a source is provided.

In our view, the objective of a first course based on this book is to help students build an intuition on how to model uncertainty into mathematical programs, which changes uncertainty brings into the decision process, what difficulties uncertainty may bring, and what problems are solvable. To begin this development, the first section in Chapter 1 provides a worked example of modeling a stochastic program. It introduces the basic concepts, without using any new or specific techniques. This first example can be complemented by any one of the other proposed cases of Chapter 1, in finance, in multistage capacity expansion, and in manufacturing. Based again on examples, Chapter 2 describes how a stochastic model is formally built. It also stresses the fact that several different models can be built, depending on the type of uncertainty and the time when decisions must be taken. This chapter links the various concepts to alternative fields of planning under uncertainty.

Any course should begin with the study of those two chapters. The sequel would then depend on the students' interests and backgrounds. A typical course would consist of elements of Chapter 3, Sections 4.1 to 4.5, Sections 5.1 to 5.3 and 5.7, and one or two more advanced sections of the instructor's choice. The final case study may serve as a conclusion. A class emphasizing modeling might focus on basic approximations in Chapter 9 and sampling in Chapter 10. A computational class would stress methods from Chapters 6 to 8. A more theoretical class might concentrate more deeply on Chapter 3 and the results from Chapters 9 to 11.

The book can also be used as an introduction for graduate students interested in stochastic programming as a research area. They will find a broad coverage of mathematical properties, models, and solution algorithms. Broad coverage cannot mean an in-depth study of all existing research. The reader will thus be referred to the original papers for details. Advanced sections may require multivariate calculus, probability measure theory, or an introduction to nonlinear or integer programming. Here again, the stress is clearly in building knowledge and intuition in the field. Mathematical results are given so long as they are either basic properties or helpful in developing efficient solution procedures. The importance of the various sections clearly reflects our own interests, which focus on results that may lead to practical applications of stochastic programming.

To conclude, we may use the following little story. An elderly person, celebrating her one hundredth birthday, was asked how she succeeded in reaching that age. She answered, "It's very simple. You just have to wait."

In comparison, stochastic programming may well look like a field of young impatient people who not only do not want to wait and see but who consider waiting to be suboptimal. We realize how much patience was needed from our friends and colleagues who encouraged us to write this book, which took us much longer than expected. To all of them, we are extremely thankful for their support. The authors also wish to thank the Fonds National de la Recherche Scientifique and the National Science Foundation for their financial support. Both authors are deeply grateful to the people who introduced us to the field, George Dantzig, Roger Wets, Jacques

Drèze, and Guy de Ghellinck. Our special thanks go to our wives, Pierrette and Marie, to whom we dedicate this book.

Ann Arbor, Michigan John R. Birge
Namur, Belgium François Louveaux

Contents

Notation

The following describes the major symbols and notations used in the text. To the greatest extent possible, we have attempted to keep unique meanings for each item. In those cases where an item has additional uses, they should be clear from context. We include here only notation used in more than one section. Additional notation may be needed within specific sections and is explained when used.

In general, vectors are assumed to be columns with transposes to indicate row vectors. This yields $c^T x$ to denote the inner product of two n-vectors, c and x. We reserve prime ($'$) for first derivatives with respect to time (e.g., $f' = df/dt$).

Vectors in primal programs are represented by lowercase Latin letters while matrices are uppercase. Dual variables and certain scalars are generally Greek letters. Superscripts indicate a stage while subscripts indicate components followed by realization index. Boldface indicates a random quantity. Expectations of random variables are indicated by a bar ($\bar{\xi}$), μ, or ($\mathrm{E}(\boldsymbol{\xi})$). We also use the bar notation to denote sample means in Chapter 9.

Equations are numbered consecutively in the text by section and number within the section (e.g., (1.2) for Section 1, Equation 2). For references to chapters other than the current one, we use three indices: chapter, section, and equation, (e.g., (3.1.2) for Chapter 3, Section 1, Equation 2). Exercises are given at the end of sections (or subsections in the cases of Sections 3.2 and 5.1) and are referenced in the same manner as equations. All other items (figures, tables, declarations, examples) are labeled consecutively through the entire chapter with a single reference (e.g., Figure 1) if within the current chapter and chapter and number if in a different chapter (e.g., Figure 3.1 for Chapter 3, Figure 1).

Symbol	Definition
$+$	Superscript indicates the positive part of a real (i.e., $a^+ = \max(a,0)$) or unrestricted variable (e.g., $y = y^+ - y^-, y^+ \geq 0, y^- \geq 0$) and its objective coefficients (e.g., q^+), subscript as non-negative values in a set (e.g., $\Re_+$) or the right-limit ($F^+(t) = \lim_{s \downarrow t} F(s)$)
$-$	Superscript indicates the negative part of a real (i.e., $a^- = \max(-a,0)$) or unrestricted variable (e.g., $y = y^+ - y^-, y^+ \geq 0, y^- \geq 0$) and its objective coefficients (e.g., q^-) or the left-limit ($F^-(t) = \lim_{s \uparrow t} F(s)$)
$*$	Indicates an optimal value or solution (e.g., x^*)
$0 \ \hat{} \ \prime \ \sim$	Indicate given nonoptimal values or solutions (e.g., x^0 , $\hat{x}$, x' , $\tilde{x}$)
0	Zero matrix (subscripts denote dimension when present)
$\mathbf{1}_X$	Indicator function of set X
a	Ancestor scenario, real value or vector
A	First-stage matrix (e.g., $Ax = b$), also used to indicate an event or subset, $A \in \mathscr{A} \subset \Omega$
$\mathscr{A}$	Collection of subsets
b	First-stage right-hand side (e.g., $Ax = b$)
B	Matrix, basis submatrix, Borel sets, or index set of a basis
$\mathscr{B}$	Collection of subsets (notably Borel sets)
c	First-stage objective ($c^T x$), t -th stage objective ($(c^t(\omega))^T x^t$) or real vectors
C	Matrix or index set of continuous variables
d	Right-hand side of a feasibility cut in the L-shaped method, a demand, or real vector
D	Left-hand side vector of a feasibility cut in the L-shaped method, a matrix, a set, or an index set of discrete variables
$\mathscr{D}$	Set of descendant scenarios
e	Exponential, right-hand side of an optimality cut in the L-shaped method, an extreme point, or the unit vector ($e^T = (1, \ldots, 1)$)
E	Mathematical expectation operator, left-hand side vector of an optimality cut in the L-shaped method, or an event
f	Function (usually in an objective ($f(x)$ or $f_i(x)$) or a density
F	Cumulative probability distribution

Symbol	Definition
g	Function (usually in constraints ($g(x)$ or $g_j(x)$))
h	Right-hand side in second-stage ($Wy = h - Tx$), also $h^t(\omega)$ in multistage problems
H	Number of stages (horizon) in multistage problems
i	Subscript index of functions (f_i) or vector elements (x_i , x_{ij})
I	Identity matrix or index set ($i \in I$)
j	Subscript index of functions (g_j) or vector elements (y_j , y_{ij})
J	Matrix or index set
k	Index of a realization of a random vector ($k = 1, \ldots, K$)
K	Feasibility sets (K_1, K_2) or total number of realizations of a discrete random vector
$\mathscr{K}$	Number of realizations or sample paths in a scenario tree with $\mathscr{K}^t$ nodes at stage t
l	Index, lower bound on a variable, or Lagrangian function
L	The L-shaped method, objective value lower bound, or real value
m	Number of constraints (m_1, m_2) or number of elements ($i = 1, \ldots, m$)
n	Number of variables (n_1, n_2) or number of elements ($i = 1, \ldots, n$)
N	Set, normal cone, normal distribution, or number of random elements
p	Probability of a random element (e.g., p_k $= \mathrm{P}(\boldsymbol{\xi} = \xi_k)$) or matrix of probabilities
P	Probability of events (e.g., $\mathrm{P}(\boldsymbol{\xi} \leq 0)$)
q	Second-stage objective vector ($q^T y$)
Q	Second-stage (multistage) value function with random argument ($Q(x, \boldsymbol{\xi})$ or $Q^t(x^t, \xi^t)$)
$\mathscr{Q}$	Second-stage (multistage) expected value value (recourse) function ($\mathscr{Q}(x)$ or $\mathscr{Q}^t(x^t)$)
r	Revenue or return in examples, real vector, or index
$\Re$	Real numbers
R	Matrix or set
s	Scenario or index

Symbol	Definition
S	Set or matrix
t	Superscript stage or period index for multistage programs ($t=1,\ldots,H$), a real-valued parameter, or an index
T	Technology matrix ($Wy=h-Tx$ or $T^{t-1}(\omega)(x)$); as a superscript, the transpose of a matrix or vector
u	General vector, upper-bound vector, or expected shortage
U	Objective value upper bound
v	Variable vector or expected surplus
V	Set, matrix or an operator
w	Second-stage decision vector in some examples
W	Recourse matrix ($Wy=h-Tx$)
x	First-stage decision vector or multistage decision vector (x^t)
X	First-stage feasible set ($x\in X$) or t th stage feasible set (X^t)
y	Second-stage decision vector
Y	Second-stage feasible set ($y\in Y$)
z	Objective value ($\min z=c^T x+\cdots$)
Z	Integers
α	Real value, vector, or probability level with probabilistic constraints
β	Real value or vector
γ	Real value or function
δ	Real value or function
ε	Real value
ζ	Random variable
η	Real value or random variable
θ	Lower bound on $\mathscr{Q}(x)$ in the L-shaped method
κ	Index
λ	Dual multiplier, parameter in a convex combination, or measure
μ	Expectation (used mostly in examples of densities) or a parameter for non-negative multiples
ν	Algorithm iteration index (sometimes also the number of samples in Monte Carlo sampling algorithms)
$\boldsymbol{\xi}$	Random vector (often indexed by time, $\boldsymbol{\xi}^t$) with realizations as ξ (without boldface)
Ξ	Support of the random vector $\boldsymbol{\xi}$
π	Dual multiplier

Symbol	Definition
Π	Product, projection operator, or aggregated problem dual multiplier
ρ	Dual multiplier or discount factor
σ	Dual multiplier, standard deviation, or σ -field
Σ	Summation or covariance matrix
τ	Possible right-hand side in bundles or index of time
ϕ	Function in computing the value of the stochastic solution or a measure
Φ	Function, cumulative distribution of standard normal
$\emptyset$	Empty set
χ	Tender or offer from first to second period ($\chi = Tx$)
ψ	Second stage value function defined on tenders and with random argument, $\psi(\chi, \xi(\omega))$)
Ψ	Expected second stage value function defined on tenders, $\Psi(\chi)$)
ω	Random event ($\omega \in \Omega$)
Ω	Set of all random events

Part I
Models

Chapter 1
Introduction and Examples

This chapter presents stochastic programming examples from a variety of areas with wide application. These examples are intended to help the reader build intuition on how to model uncertainty. They also reflect different structural aspects of the problems. In particular, we show the variety of stochastic programming models in terms of the objectives of the decision process, the constraints on those decisions, and their relationships to the random elements.

In each example, we investigate the value of the stochastic programming model over a similar deterministic problem. We show that even simple models can lead to significant savings. These results provide the motivation to lead us into the following chapters on stochastic programs, solution properties, and techniques.

In the first section, we consider a farmer who must decide on the amounts of various crops to plant. The yields of the crops vary according to the weather. From this example, we illustrate the basic foundation of stochastic programming and the advantage of the stochastic programming solution over deterministic approaches. We also introduce the classical news vendor (or newsboy) problem and give the fundamental properties of these problems' general class, called *two-stage stochastic linear programs with recourse*.

The second section contains an example in planning finances for a child's education. This example fits the situation in many discrete time control problems. Decisions occur at different points in time so that the problem can be viewed as having multiple stages of observations and actions.

The third section considers power system capacity expansion. Here, decisions are taken dynamically about additional capacity and about the allocation of capacity to meet demand. The resulting problem has multiple decision stages and a valuable property known as *block separable recourse* that allows efficient solution. The problem also provides a natural example of constraints on reliability within the area called *probabilistic* or *chance-constrained programming*.

The fourth example concerns the design of a simple axle. It includes market reaction to the design and performance characteristics of products made by a manufacturing system with variable performance. The essential characteristics of the

J.R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-1-4614-0237-4_1,
© Springer Science+Business Media, LLC 2011

maximum performance of the product illustrate a problem with fundamental nonlinearities incorporated directly into the stochastic program.

The fifth section presents a simple routing problem. It illustrates models where some decisions (traveling on an arc or not) are represented by integer decision variables. As this example is easily illustrated and does not require any solver, it may also be used as a preliminary example.

The final section of this chapter briefly describes several other major application areas of stochastic programs. The exercises at the end of the chapter develop modeling techniques. This chapter illustrates some of the range of stochastic programming applications but is not meant to be exhaustive. Applications in location and distribution, for example, are discussed in Chapter 2.

1.1 A Farming Example and the News Vendor Problem

a. The farmer's problem

Consider a European farmer who specializes in raising wheat, corn, and sugar beets on his 500 acres of land. During the winter, he wants to decide how much land to devote to each crop. (We refer to the farmer as "he" for convenience and not to imply anything about the gender of European farmers.)

The farmer knows that at least 200 tons (T) of wheat and 240 T of corn are needed for cattle feed. These amounts can be raised on the farm or bought from a wholesaler. Any production in excess of the feeding requirement would be sold. Over the last decade, mean selling prices have been \$170 and \$150 per ton of wheat and corn, respectively. The purchase prices are 40% more than this due to the wholesaler's margin and transportation costs.

Another profitable crop is sugar beet, which he expects to sell at \$36/T; however, the European Commission imposes a quota on sugar beet production. Any amount in excess of the quota can be sold only at \$10/T. The farmer's quota for next year is 6000 T.

Based on past experience, the farmer knows that the mean yield on his land is roughly 2.5 T, 3 T, and 20 T per acre for wheat, corn, and sugar beets, respectively. Table 1 summarizes these data and the planting costs for these crops.

To help the farmer make up his mind, we can set up the following model. Let

x_1 = acres of land devoted to wheat,
x_2 = acres of land devoted to corn,
x_3 = acres of land devoted to sugar beets,
w_1 = tons of wheat sold,
y_1 = tons of wheat purchased,
w_2 = tons of corn sold,
y_2 = tons of corn purchased,
w_3 = tons of sugar beets sold at the favorable price,

Table 1 Data for farmer's problem.

	Wheat	Corn	Sugar Beets
Yield (T/acre)	2.5	3	20
Planting cost ($/acre)	150	230	260
Selling price ($/T)	170	150	36 under 6000 T 10 above 6000 T
Purchase price ($/T)	238	210	–
Minimum requirement (T)	200	240	–
Total available land: 500 acres			

73 w_4 = tons of sugar beets sold at the lower price.

The problem reads as follows:

$$
\begin{aligned}
\min\ & 150x_1 + 230x_2 + 260x_3 + 238y_1 - 170w_1 \\
& + 210y_2 - 150w_2 - 36w_3 - 10w_4 \\
\text{s. t.}\ & x_1 + x_2 + x_3 \leq 500\ ,\ 2.5\,x_1 + y_1 - w_1 \geq 200\ , \\
& 3\,x_2 + y_2 - w_2 \geq 240\ ,\ w_3 + w_4 \leq 20x_3, w_3 \leq 6000\ , \\
& x_1, x_2, x_3,\ y_1, y_2,\ w_1, w_2, w_3, w_4 \geq 0\ .
\end{aligned}
\tag{1.1}
$$

After solving (1.1) with his favorite linear program solver, the farmer obtains an optimal solution, as in Table 2.

Table 2 Optimal solution based on expected yields.

Culture	Wheat	Corn	Sugar Beets
Surface (acres)	120	80	300
Yield (T)	300	240	6000
Sales (T)	100	–	6000
Purchase (T)	–	–	–
Overall profit: $118,600			

This optimal solution is easy to understand. The farmer devotes enough land to sugar beets to reach the quota of 6000 T. He then devotes enough land to wheat and corn production to meet the feeding requirement. The rest of the land is devoted to wheat production. Some wheat can be sold.

To an extent, the optimal solution follows a very simple heuristic rule: to allocate land in order of decreasing profit per acre. In this example, the order is sugar beets at a favorable price, wheat, corn, and sugar beets at the lower price. This simple heuristic would, however, no longer be valid if other constraints, such as labor requirements or crop rotation, would be included.

After thinking about this solution, the farmer becomes worried. He has indeed experienced quite different yields for the same crop over different years mainly because of changing weather conditions. Most crops need rain during the few weeks after seeding or planting, then sunshine is welcome for the rest of the growing period. Sunshine should, however, not turn into drought, which causes severe yield reductions. Dry weather is again beneficial during harvest. From all these factors, yields varying 20 to 25% above or below the mean yield are not unusual.

In the next sections, we study two possible representations of these variable yields. One approach using discrete, correlated random variables is described in Sections 1.1b. and 1.1c. Another, using continuous uncorrelated random variables, is described in Section 1.1d.

The influence of price fluctuations, illustrated by the dramatic price increases in 2007, is discussed in Exercise 8.

b. A scenario representation

A first possibility is to assume some correlation among the yields of the different crops. A very simplified representation of this would be to assume that years are good, fair, or bad for all crops, resulting in above average, average, or below average yields for all crops. To fix these ideas, "above" and "below" average indicate a yield 20% above or below the mean yield given in Table 1. For simplicity, we assume that weather conditions and yields for the farmer do not have a significant impact on prices.

The farmer wishes to know whether the optimal solution is sensitive to variations in yields. He decides to run two more optimizations based on above average and below average yields. Tables 3 and 4 give the optimal solutions he obtains in these cases.

Again, the solutions in Tables 3 and 4 seem quite natural. When yields are high, smaller surfaces are needed to raise the minimum requirements in wheat and corn and the sugar beet quota. The remaining land is devoted to wheat, whose extra production is sold. When yields are low, larger surfaces are needed to raise the minimum requirements and the sugar beet quota. In fact, corn requirements cannot be satisfied with the production, and some corn must be bought.

The optimal solution is very sensitive to changes in yields. The optimal surfaces devoted to wheat range from 100 acres to 183.33 acres. Those devoted to corn range from 25 acres to 80 acres and those devoted to sugar beets from 250 acres to 375 acres. The overall profit ranges from $59,950 to $167,667.

Long-term weather forecasts would be very helpful here. Unfortunately, as even meteorologists agree, weather conditions cannot be accurately predicted six months ahead. The farmer must make up his mind without perfect information on yields.

Table 3 Optimal solution based on above average yields (+ 20%).

Culture	Wheat	Corn	Sugar Beets
Surface (acres)	183.33	66.67	250
Yield (T)	550	240	6000
Sales (T)	350	–	6000
Purchase (T)	–	–	–
Overall profit: $167,667			

Table 4 Optimal solution based on below average yields (-20%).

Culture	Wheat	Corn	Sugar Beets
Surface (acres)	100	25	375
Yield (T)	200	60	6000
Sales (T)	–	–	6000
Purchase (T)	–	180	–
Overall profit: $59,950			

The main issue here is clearly on sugar beet production. Planting large surfaces would make it certain to produce and sell the quota, but would also make it likely to sell some sugar beets at the unfavorable price. Planting small surfaces would make it likely to miss the opportunity to sell the full quota at the favorable price.

The farmer now realizes that he is unable to make a perfect decision that would be best in all circumstances. He would, therefore, want to assess the benefits and losses of each decision in each situation. Decisions on land assignment (x_1,x_2,x_3) have to be taken now, but sales and purchases $(w_i,\ i=1,\ldots,4,\ y_j,\ j=1,2)$ depend on the yields. It is useful to index those decisions by a scenario index $s=1,2,3$ corresponding to above average, average, or below average yields, respectively. This creates a new set of variables of the form w_{is} , $i=1,2,3,4$, $s=1,2,3$ and y_{js} , $j=1,2$, $s=1,2,3$. As an example, w_{32} represents the amount of sugar beets sold at the favorable price if yields are average.

Assuming the farmer wants to maximize long-run profit, it is reasonable for him to seek a solution that maximizes his expected profit. (This assumption means that the farmer is neutral about risk. For a discussion of risk aversion and alternative utilities, see Chapter 2.) If the three scenarios have an equal probability of $1/3$, the farmer's problem reads as follows:

$$
\begin{aligned}
\min\ & 150x_1 + 230x_2 + 260x_3 \\
& \quad - \tfrac{1}{3}(170w_{11} - 238y_{11} + 150w_{21} - 210y_{21} + 36w_{31} + 10w_{41}) \\
& \quad - \tfrac{1}{3}(170w_{12} - 238y_{12} + 150w_{22} - 210y_{22} + 36w_{32} + 10w_{42}) \\
& \quad - \tfrac{1}{3}(170w_{13} - 238y_{13} + 150w_{23} - 210y_{23} + 36w_{33} + 10w_{43}) \\
\text{s.t. } & x_1 + x_2 + x_3 \le 500\ ,\ 3x_1 + y_{11} - w_{11} \ge 200\ , \\
& 3.6x_2 + y_{21} - w_{21} \ge 240\ ,\ w_{31} + w_{41} \le 24x_3\ ,\ w_{31} \le 6000\ , \\
& 2.5x_1 + y_{12} - w_{12} \ge 200\ ,\ 3x_2 + y_{22} - w_{22} \ge 240\ , \\
& w_{32} + w_{42} \le 20x_3\ ,\ w_{32} \le 6000\ ,\ 2x_1 + y_{13} - w_{13} \ge 200, \\
& 2.4x_2 + y_{23} - w_{23} \ge 240,\ w_{33} + w_{43} \le 16x_3\ , \\
& w_{33} \le 6000,\ x, y, w \ge 0\ .
\end{aligned}
\tag{1.2}
$$

Such a model of a stochastic decision program is known as the *extensive form* of the stochastic program because it explicitly describes the second-stage decision variables for all scenarios. The optimal solution of (1.2) is given in Table 5. The top line gives the planting areas, which must be determined before realizing the weather and crop yields. This decision is called the *first stage*. The other lines describe the yields, sales, and purchases in the three scenarios. They are called the *second stage*. The bottom line shows the overall expected profit.

Table 5 Optimal solution based on the stochastic model (1.2).

		Wheat	Corn	Sugar Beets
First Stage	Area (acres)	170	80	250
$s = 1$ Above	Yield (T)	510	288	6000
	Sales (T)	310	48	6000 (favor. price)
	Purchase (T)	–	–	–
$s = 2$ Average	Yield (T)	425	240	5000
	Sales (T)	225	–	5000 (favor. price)
	Purchase (T)	–	–	–
$s = 3$ Below	Yield (T)	340	192	4000
	Sales (T)	140	–	4000 (favor. price)
	Purchase (T)	–	48	–
Overall profit: \$108,390				

The optimal solution can be understood as follows. The most profitable decision for sugar beet land allocation is the one that always avoids sales at the unfavorable price even if this implies that some portion of the quota is unused when yields are average or below average.

The area devoted to corn is such that it meets the feeding requirement when yields are average. This implies sales are possible when yields are above average and purchases are needed when yields are below average. Finally, the rest of the land is devoted to wheat. This area is large enough to cover the minimum requirement. Sales then always occur.

This solution illustrates that it is impossible, under uncertainty, to find a solution that is ideal under all circumstances. Selling some sugar beets at the unfavorable price or having some unused quota is a decision that would never take place with a perfect forecast. Such decisions can appear in a stochastic model because decisions have to be balanced or hedged against the various scenarios.

The hedging effect has an important impact on the expected optimal profit. Suppose yields vary over years but are cyclical. A year with above average yields is always followed by a year with average yields and then a year with below average yields. The farmer would then take optimal solutions as given in Table 3, then Table 2, then Table 4, respectively. This would leave him with a profit of \$167,667 the first year, \$118,600 the second year, and \$59,950 the third year. The mean profit over the three years (and in the long run) would be the mean of the three figures, namely \$115,406 per year.

Now, assume again that yields vary over years, but on a random basis. If the farmer gets the information on the yields before planting, he will again choose the areas on the basis of the solution in Table 2, 3, or 4, depending on the information received. In the long run, if each yield is realized one third of the years, the farmer will get again an expected profit of \$115,406 per year. This is the situation under perfect information.

As we know, the farmer unfortunately does not get prior information on the yields. So, the best he can do in the long run is to take the solution as given by Table 5. This leaves the farmer with an expected profit of \$108,390. The difference between this figure and the value, \$115,406, in the case of perfect information, namely \$7016, represents what is called *the expected value of perfect information* (*EVPI*). This concept, along with others, will be studied in Chapter 4. At this introductory level, we may just say that it represents the loss of profit due to the presence of uncertainty.

Another approach the farmer may have is to assume expected yields and always to allocate the optimal planting surface according to these yields, as in Table 2. This approach represents the *expected value solution*. It is common in optimization but can have unfavorable consequences. Here, as shown in Exercise 1, using the expected value solution every year results in a long run annual profit of \$107,240. The loss by not considering the random variations is the difference between this and the stochastic model profit from Table 5. This value, \$108,390 − 107,240=\$1,150, is the *value of the stochastic solution* (*VSS*), the possible gain from solving the stochastic model. Note that it is not equal to the expected value of perfect information, and, as we shall see in later models, may in fact be larger than the *EVPI* .

These two quantities give the motivation for stochastic programming in general and remain a key focus throughout this book. *EVPI* measures the value of knowing the future with certainty while *VSS* assesses the value of knowing and using distributions on future outcomes. Our emphasis will be on problems where no further information about the future is available so the *VSS* becomes more practically relevant. In some situations, however, more information might be available through more extensive forecasting, sampling, or exploration. In these cases, $EVPI$ would be useful for deciding whether to undertake additional efforts.

c. General model formulation

We may also use this example to illustrate the general formulation of a stochastic problem. We have a set of decisions to be taken without full information on some random events. These decisions are called *first-stage decisions* and are usually represented by a vector x. In the farmer example, they are the decisions on how many acres to devote to each crop. Later, full information is received on the realization of some random vector $\boldsymbol{\xi}$. Then, second-stage or corrective actions $\mathbf{y}$ are taken. We use boldface notation here and throughout the book to denote that these vectors are random and to differentiate them from their realizations. We also sometimes use a functional form, such as $\xi(\omega)$ or $y(s)$, to show explicit dependence on an underlying element, ω or s.

In the farmer example, the random vector is the set of yields and the corrective actions are purchases and sales of products. In mathematical programming terms, this defines the so-called two-stage stochastic program with recourse of the form

$$\begin{aligned} \min\ & c^T x + \mathrm{E}_{\boldsymbol{\xi}} Q(x,\boldsymbol{\xi}) \\ \text{s. t. } & Ax = b\ , \\ & x \geq 0\ , \end{aligned} \tag{1.3}$$

where $Q(x,\boldsymbol{\xi}) = \min\{\mathbf{q}^T\mathbf{y} \mid W\mathbf{y} = \mathbf{h} - \mathbf{T}x, y \geq 0\}$, $\boldsymbol{\xi}$ is the vector formed by the components of $\mathbf{q}^T$, $\mathbf{h}^T$, and $\mathbf{T}$, and $\mathrm{E}_{\boldsymbol{\xi}}$ denote mathematical expectation with respect to $\boldsymbol{\xi}$. We assume here that W is fixed (*fixed recourse*). Reasons for this restriction are explained in Section 3.1.

In the farmer example, the random vector is a discrete variable with only three different values. Only the T matrix is random. A second-stage problem for one particular scenario s can thus be written as

$$\begin{aligned} Q(x,s) = \min\ & \{238y_1 - 170w_1 + 210y_2 - 150w_2 - 36w_3 - 10w_4\} \\ \text{s. t. } & t_1(s)x_1 + y_1 - w_1 \geq 200\ , \\ & t_2(s)x_2 + y_2 - w_2 \geq 240\ , \\ & w_3 + w_4 \leq t_3(s)x_3\ , \\ & w_3 \leq 6000\ , \\ & y, w \geq 0\ , \end{aligned} \tag{1.4}$$

where $t_i(s)$ represents the yield of crop i under scenario s (or state of nature s). To illustrate the link between the general formulation (1.3) and the example (1.4), observe that in (1.4) we may say that the random vector $\boldsymbol{\xi} = (\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3)$ is formed by the three yields and that $\boldsymbol{\xi}$ can take on three different values, say ξ_1, ξ_2, and ξ_3, which represent $(t_1(1), t_2(1), t_3(1))$, $(t_1(2), t_2(2), t_3(2))$, and $(t_1(3), t_2(3), t_3(3))$, respectively.

An alternative interpretation would be to say that the random vector $\xi(s)$ in fact depends on the scenario s, which takes on three different values[1].

In this section, we have illustrated two possible representations of a stochastic program. The form (1.2) given earlier for the farmer's example is known as the extensive form. It is obtained by associating one decision vector in the second-stage to each possible realization of the random vector. The second form (1.3) or (1.4) is called the implicit representation of the stochastic program. A more condensed implicit representation is obtained by defining $\mathscr{Q}(x) = \mathrm{E}_{\boldsymbol{\xi}} Q(x, \boldsymbol{\xi})$ as the *value function* or *recourse function* so that (1.3) can be written as

$$\begin{aligned} \min\ & c^T x + \mathscr{Q}(x) \\ \text{s. t. } & Ax = b\,, \\ & x \geq 0\,. \end{aligned} \tag{1.5}$$

d. Continuous random variables

Contrary to the assumption made in Section 1.1b., we may also assume that yields for the different crops are independent. In that case, we may as well consider a continuous random vector for the yields. To illustrate this, let us assume that the yield for each crop i can be appropriately described by a uniform random variable, inside some range $[l_i, u_i]$ (see Appendix A.2). For the sake of comparison, we may take l_i to be 80% of the mean yield and u_i to be 120% of the mean yield so that the expectations for the yields will be the same as in Section 1.1b. Again, the decisions on land allocation are first-stage decisions because they are taken before knowledge of the yields. Second-stage decisions are purchases and sales after the growing period. The second-stage formulation can again be described as $\mathscr{Q}(x) = \mathrm{E}_{\boldsymbol{\xi}} Q(x, \boldsymbol{\xi})$, where $Q(x, \boldsymbol{\xi})$ is the value of the second stage for a given realization of the random vector.

Now, in this particular example, the computation of $Q(x, \boldsymbol{\xi})$ can be separated among the three crops due to independence of the random vector. (Note that this separability property also holds in the discrete representation of Section 1.1b.) We can then write:

$$\mathrm{E}_{\boldsymbol{\xi}} Q(x, \boldsymbol{\xi}) = \sum_{i=1}^{3} \mathrm{E}_{\boldsymbol{\xi}} Q_i(x_i, \boldsymbol{\xi}) = \sum_{i=1}^{3} \mathscr{Q}_i(x_i)\,, \tag{1.6}$$

where $Q_i(x_i, \boldsymbol{\xi})$ is the optimal second-stage value of purchases and sales of crop i.

We are in fact in position to give an exact analytical expression for the second-stage value functions $\mathscr{Q}_i(x_i)$, $i = 1, \ldots, 3$. We first consider sugar beet sales. For

[1] Note that the decisions y_1, y_2, w_1, w_2, w_3, and w_4 also depend on the scenario. This dependence is not always made explicit. It appears explicitly in (1.7) but not in (1.4).

a given value $t_3(\xi)$ of the sugar beet yield, one obtains the following second-stage problem:

$$\begin{aligned} Q_3(x_3,\xi) = \min\ & -36w_3(\xi) - 10w_4(\xi) \\ \text{s. t. }\ & w_3(\xi) + w_4(\xi) \le t_3(\xi)x_3 , \\ & w_3(\xi) \le 6000 , \\ & w_3(\xi), w_4(\xi) \ge 0 . \end{aligned} \tag{1.7}$$

The optimal decisions for this problem are clearly to sell as many sugar beets as possible at the favorable price, and to sell the possible remaining production at the unfavorable price, namely

$$\begin{aligned} w_3(\xi) &= \min[6000, t_3(\xi)x_3] , \\ w_4(\xi) &= \max[t_3(\xi)x_3 - 6000, 0] . \end{aligned} \tag{1.8}$$

This results in a second-stage value of

$$Q_3(x_3,\xi) = -36\min[6000, t_3(\xi)x_3] - 10\max[t_3(\xi)x_3 - 6000, 0] .$$

We first assume that the surface x_3 devoted to sugar beets will not be so large that the quota would be exceeded for any possible yield or so small that production would always be less than the quota for any possible yield. In other words, we assume that the following relation holds:

$$l_3x_3 \le 6000 \le u_3x_3 , \tag{1.9}$$

where, as already defined, l_3 and u_3 are the bounds on the possible values of $t_3(\xi)$. Under this assumption, the expected value of the second stage for sugar beet sales is

$$\begin{aligned} \mathscr{Q}_3(x_3) &= \mathrm{E}_{\xi} Q_3(x_3,\xi_3) \\ &= -\int_{l_3}^{6000/x_3} 36tx_3 f(t)dt \\ &\quad - \int_{6000/x_3}^{u_3} (216000 + 10tx_3 - 60000) f(t)dt, \end{aligned}$$

where $f(t)$ denotes the density of the random yield $t_3(\xi)$. Given the assumption that this density is uniform over the interval $[l_3, u_3]$, one obtains, after some computation, the following analytical expression

$$\mathscr{Q}_3(x_3) = -18\frac{(u_3^2 - l_3^2)x_3}{u_3 - l_3} + \frac{13(u_3x_3 - 6000)^2}{x_3(u_3 - l_3)} ,$$

which can also be expressed as

$$\mathscr{Q}_3(x_3) = -36\bar{t}_3x_3 + \frac{13(u_3x_3 - 6000)^2}{x_3(u_3 - l_3)} , \tag{1.10}$$

where $\bar{t}_3$ denotes the expected yield for sugar beet production, which is $\frac{u_3+l_3}{2}$ for a uniform density.

Note that assumption (1.9) is not really limiting. We can still compute the analytical expression of $\mathscr{Q}_3(x_3)$ for the other situations.

For example, if the surface x_3 is such that the production exceeds the quota for any possible yield $(l_3 x_3 > 6000)$, then the optimal second-stage decisions are simply

$$w_3(\boldsymbol{\xi}) = 6000 \,,$$
$$w_4(\boldsymbol{\xi}) = t_3(\boldsymbol{\xi})x_3 - 6000 \text{ ,for all } \boldsymbol{\xi} \,.$$

The second-stage value for a given ξ is now

$$Q_3(x_3,\xi) = -216000 - 10(t_3(\xi)x_3 - 6000) = -156000 - 10t_3(\xi)x_3 \,,$$

and the expected value is simply

$$\mathscr{Q}_3(x_3) = -156000 - 10\bar{t}_3 x_3 \,. \tag{1.11}$$

Similarly, if the surface devoted to sugar beets is so small that for any yield the production is lower than the quota, the second-stage value function is

$$\mathscr{Q}_3(x_3) = -36\bar{t}_3 x_3 \,. \tag{1.12}$$

We may therefore draw the graph of the function $\mathscr{Q}_3(x_3)$ for all possible values of x_3 as in Figure 1. Note that with our assumption of $\bar{t}_3 = 20$, we would then have the limits on x_3 in (1.9) as $250 \le x_3 \le 375$.

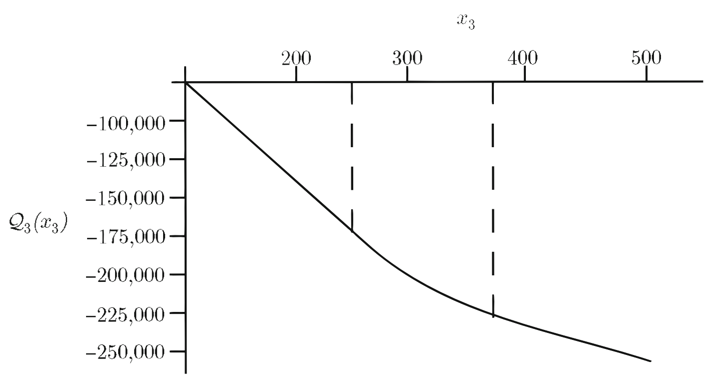

Fig. 1 The expected recourse value for sugar beets as a function of acres planted.

We immediately see that the function has three different pieces. Two of these pieces are linear and one is nonlinear, but the function $\mathscr{Q}_3(x_3)$ is continuous and convex. This property will be proved when we consider the generalization of this problem, known as the *news vendor, newsboy, or Christmas tree problem*. In fact, this property holds for a large class of second-stage problems, as will be seen in Chapter 3.

Similar computations can be done for the other two crops. For wheat, we obtain

$$\mathscr{Q}_1(x_1) = \begin{cases} 47600 - 595x_1 & \text{for } x_1 \leq 200/3\ , \\ 119\frac{(200-2x_1)^2}{x_1} - 85\frac{(200-3x_1)^2}{x_1} & \text{for } \frac{200}{3} \leq x_1 \leq 100\ , \\ 34000 - 425x_1 & \text{for } x_1 \geq 100\ , \end{cases}$$

and, for corn, we obtain

$$\mathscr{Q}_2(x_2) = \begin{cases} 50400 - 630x_2 & \text{for } x_2 \leq 200/3\ , \\ 87.5\frac{(240-2.4x_2)^2}{x_2} - 62.5\frac{(240-3.6x_2)^2}{x_2} & \text{for } 200/3 \leq x_2 \leq 100\ , \\ 36000 - 450x_2 & \text{for } x_2 \geq 100\ . \end{cases}$$

The global problem is therefore

$$\begin{aligned} &\min\ 150x_1 + 230x_2 + 260x_3 + \mathscr{Q}_1(x_1) + \mathscr{Q}_2(x_2) + \mathscr{Q}_3(x_3) \\ &\text{s. t. } x_1 + x_2 + x_3 \leq 500\ , \\ &\qquad x_1, x_2, x_3 \geq 0\ . \end{aligned}$$

Given that the three functions $\mathscr{Q}_i(x_i)$ are convex, continuous, and differentiable functions and the first-stage objective is linear, this problem is a convex program for which Karush-Kuhn-Tucker (K-K-T) conditions are necessary and sufficient for a global optimum. (This result is from nonlinear programming. For more on this result about optimality, see Section 2.11.) Denoting by λ the multiplier of the surface constraint and as before by c_i the first-stage objective coefficient of crop i, the K-K-T conditions require

$$\begin{aligned} &x_i\left[c_i + \frac{\partial \mathscr{Q}_i(x_i)}{\partial x_i} + \lambda\right] = 0\ ,\ \ c_i + \frac{\partial \mathscr{Q}_i(x_i)}{\partial x_i} + \lambda \geq 0\ ,\ x_i \geq 0\ ,\ i = 1,2,3\ ; \\ &\lambda[x_1 + x_2 + x_3 - 500] = 0\ ,\quad x_1 + x_2 + x_3 \leq 500\ ,\ \ \lambda \geq 0\ . \end{aligned}$$

Assume the optimal solution is such that $100 \leq x_1$, $\frac{200}{3} \leq x_2 \leq 100$, and $250 \leq x_3 \leq 375$ with $\lambda \neq 0$. Then the conditions read

$$\begin{cases} -275 + \lambda = 0 \ , \\ -76 - \frac{1.44\ 10^6}{x_2^2} + \lambda = 0 \ , \\ 476 - \frac{5.85\ 10^7}{x_3^2} + \lambda = 0 \ , \\ x_1 + x_2 + x_3 = 500 \ . \end{cases}$$

Solving this system of equations gives $\lambda = 275.00$, $x_1 = 135.83$, $x_2 = 85.07$, $x_3 = 279.10$, which satisfies all the required conditions and is therefore optimal. We observe that this solution is similar to the one obtained by using the scenario approach, although more surface is devoted to sugar beet and less to wheat than before. This similarity represents a characteristic robustness of a well-formed stochastic programming formulation. We shall consider it in more detail in our discussion of approximations in Chapter 8.

e. The news vendor problem

The previous section illustrates an example of a famous and basic problem in stochastic optimization, *the news vendor problem*. In this problem, a news vendor goes to the publisher every morning and buys x newspapers at a price of c per paper. This number is usually bounded above by some limit u, representing either the news vendor's purchase power or a limit set by the publisher to each vendor. The vendor then walks along the streets to sell as many newspapers as possible at the selling price q. Any unsold newspaper can be returned to the publisher at a return price r, with $r < c$.

We are asked to help the news vendor decide how many newspapers to buy every morning. Demand for newspapers varies over days and is described by a random variable $\boldsymbol{\xi}$.

It is assumed here that the news vendor cannot return to the publisher during the day to buy more newspapers. Other news vendors would have taken the remaining newspapers. Readers also only want the last edition.

To describe the news vendor's profit, we define y as the effective sales and w as the number of newspapers returned to the publisher at the end of the day. We may then formulate the problem as

$$\begin{aligned} \min\ & cx + \mathcal{Q}(x) \\ & 0 \le x \le u \ , \end{aligned}$$

where

$$\mathcal{Q}(x) = \mathrm{E}_{\boldsymbol{\xi}} Q(x, \boldsymbol{\xi})$$

and

$$\begin{aligned} Q(x,\boldsymbol{\xi}) = \min \ & -qy(\boldsymbol{\xi}) - rw(\boldsymbol{\xi}) \\ \text{s. t.} \ & y(\boldsymbol{\xi}) \le \boldsymbol{\xi} \ , \\ & y(\boldsymbol{\xi}) + w(\boldsymbol{\xi}) \le x \ , \\ & y(\boldsymbol{\xi}), w(\boldsymbol{\xi}) \ge 0 \ , \end{aligned}$$

where again $\mathrm{E}_{\boldsymbol{\xi}}$ denotes the mathematical expectation with respect to $\boldsymbol{\xi}$.

In this notation, $-\mathscr{Q}(x)$ is the expected profit on sales and returns, while $-Q(x,\boldsymbol{\xi})$ is the profit on sales and returns if the demand is at level $\boldsymbol{\xi}$. The model illustrates the two-stage aspect of the news vendor problem. The buying decision has to be taken before any information is given on the demand. When demand is known in the so-called second stage, which represents the end of the sales period of a given edition, the profit can be computed. This is done using the following simple rule:

$$\begin{aligned} y^*(\boldsymbol{\xi}) \ &= \min(\boldsymbol{\xi}, x) \ , \\ w^*(\boldsymbol{\xi}) \ &= \max(x - \boldsymbol{\xi}, 0) \ . \end{aligned}$$

Sales can never exceed the number of available newspapers or the demand. Returns occur only when demand is less than the number of newspapers available. The second-stage expected value function is simply

$$\mathscr{Q}(x) = \mathrm{E}_{\boldsymbol{\xi}}[-q\min(\boldsymbol{\xi}, x) - r\max(x - \boldsymbol{\xi}, 0)] \ .$$

As we will learn later, this function is convex and continuous. It is also differentiable when $\boldsymbol{\xi}$ is a continuous random vector. In that case, the optimal solution of the news vendor's problem is simply:

$$\begin{cases} x = 0 & \text{if } c + \mathscr{Q}'(0) > 0 \ , \\ x = u & \text{if } c + \mathscr{Q}'(u) < 0 \ , \\ \text{a solution of } c + \mathscr{Q}'(x) = 0 & \text{otherwise,} \end{cases}$$

where $\mathscr{Q}'(x)$ denotes the first order derivative of $\mathscr{Q}(x)$ evaluated at x.

By construction, $\mathscr{Q}(x)$ can be computed as

$$\begin{aligned} \mathscr{Q}(x) \ &= \int_{-\infty}^{x} (-q\xi - r(x - \xi)) dF(\xi) + \int_{x}^{\infty} -qx \, dF(\xi) \\ &= -(q - r) \int_{-\infty}^{x} \xi \, dF(\xi) - rx \, F(x) - qx(1 - F(x)) \ , \end{aligned}$$

where $F(\boldsymbol{\xi})$ represents the cumulative probability distribution of $\boldsymbol{\xi}$ (see Section 2.1).

Integrating by parts, we observe that

$$\int_{-\infty}^{x} \xi \, dF(\xi) = xF(x) - \int_{-\infty}^{x} F(\xi) d\xi$$

under mild conditions on the distribution function $F(\boldsymbol{\xi})$. It follows that

$$\mathcal{Q}(x) = -qx + (q-r)\int_{-\infty}^{x} F(\xi)d\xi \ .$$

We may thus conclude that

$$\mathcal{Q}'(x) = -q + (q-r)F(x)$$

and therefore that the optimal solution is

$$\begin{cases} x^* = 0 & \text{if } \frac{q-c}{q-r} < F(0) \ , \\ x^* = u & \text{if } \frac{q-c}{q-r} > F(u) \ , \\ x^* = F^{-1}(\frac{q-c}{q-r}) & \text{otherwise,} \end{cases}$$

where $F^{-1}(\alpha)$ is the α *-quantile* of F (see Section 2.1). If F is continuous, $x = F^{-1}(\alpha)$ means $\alpha = F(x)$. Any reasonable representation of the demand would imply $F(0) = 0$ so that the solution is never $x^* = 0$.

As we shall see in Chapter 3, this problem is an example of a basic type of stochastic program called the *stochastic program with simple recourse*. The ideas of this section can be generalized to larger problems in this class of examples. Also observe that, as such, we only come to a partial answer, under the form of an expression for x^*. The vendor may still need to consult a statistician, who would provide an accurate cumulative distribution $F(\cdot)$. Only then will a precise figure be available for x^*.

Exercises

1. *Value of the stochastic solution*
   Assume the farmer allocates his land according to the solution of Table 2, i.e., 120 acres for wheat, 80 acres for corn, and 300 acres for sugar beets. Show that if yields are random (20% below average, average, and 20% above average for all crops with equal probability one third), his expected annual profit is \$107,240. To do this observe that planting costs are certain but sales and purchases depend on the yield. In other words, fill in a table such as Table 5 but with the first-stage decisions given here.

2. *Price effect*
   When yields are good for the farmer, they are usually also good for many other farmers. The supply is thus increasing, which will lower the prices. As an example, we may consider prices going down by 10% for corn and wheat when yields are above average and going up by 10% when yields are below average. Formulate the model where these changes in prices affect both sales and purchases of corn and wheat. Assume sugar beet prices are not affected by yields.

3. *Binary first stage*
   Consider the case where the farmer possesses four fields of sizes 185, 145, 105, and 65 acres, respectively. Observe that the total of 500 acres is unchanged. Now, the fields are unfortunately located in different parts of the village. For reasons of efficiency the farmer wants to raise only one type of crop on each field. Formulate this model as a two-stage stochastic program with a first-stage program with binary variables.

4. *Integer second stage*
   Consider the case where sales and purchases of corn and wheat can only be obtained through contracts involving multiples of hundred tons. Formulate the model as a stochastic program with a mixed-integer second stage.

5. Consider any one of Exercises 2 to 4. Using standard mixed integer programming software, obtain an optimal solution of the extensive form of the stochastic program. Compute the expected value of perfect information and the value of the stochastic solution.

6. *Multistage program*
   It is typical in farming to implement crop rotation in order to maintain good soil quality. Sugar beets would, for example, appear in triennial crop rotation, which means they are planted on a given field only one out of three years. Formulate a multistage program to describe this situation. To keep things simple, describe the case when sugar beets cannot be planted two successive years on the same field, and assume no such rule applies for wheat and corn.

   (On a two-year basis, this exercise consists purely of formulation: with the basic data of the example, the solution is clearly to repeat the optimal solution in Table 5, i.e., to plant 170 acres of wheat, 80 acres of corn, and 250 acres of sugar beets. The problem becomes more relevant on a three-year basis. It is also relevant on a two-year basis with fields of the sizes given in Exercise 1.

   In terms of formulation, it is sufficient to consider a three-stage model. The first stage consists of first-year planting. The second stage consists of first-year purchases and sales and second-year planting. The third-stage consists of second-year purchases and sales. Alternatively, a four-stage model can be built, separating first-year purchases and sales from second-year planting. Also discuss the question of discounting the revenues and expenses of the various stages.)

7. *Risk aversion*
   Economic theory tells us that, like many other people, the farmer would normally act as a risk-averse person. There are various ways to model risk aversion. One simple way is to plan for the worst case. More precisely, it consists of maximizing the profit under the worst situation. Note that for some models, it is not known in advance which scenario will turn out to induce the lowest profit.

In our example, the worst situation corresponds to Scenario 3 (below average yields). Planning for the worst case implies the solution of Table 4 is optimal.

(a) Compute the loss in expected profit if that solution is taken.
(b) A median situation would be to require a reasonable profit under the worst case. Find the solution that maximizes the expected profit under the constraint that in the worst case the profit does not fall below \$58,000. What is now the loss in expected profit?
(c) Repeat part (b) with other values of minimal profit: \$56,000, \$54,000, \$52,000, \$50,000, and \$48,000. Graph the curve of expected profit loss. Also compare the associated optimal decisions.

8. *Data fluctuations*
Table 1 contains mean data over a relatively long period, from the late nineties till 2006. Yield fluctuations have been treated through random yields. What about other data's fluctuations? Planting costs in euros have not changed so much over time. (The story is different when expressed in dollars. However, the farmer's decisions are unaffected by currency modifications as they simply shift the objective function. The only element which could be affected by currency rates is the world price of sugar beets, but it has stayed low enough to play no significant role for the farmer.) Starting from the deterministic model (1.1), sensitivity analysis tells us that the optimal solution remains valid if wheat and corn selling prices remain below 220 and 168.333, respectively, and if sugar beet's favorable price remains over 26.75. This implies the solution of model (1.1) remains stable even if relatively large changes in prices occur (with the provision that the results of linear programming sensitivity analysis are guaranteed to hold when only one price is changing at a time). For joint modifications of prices, it is interesting to look at the returns of each crop. Then, one can see that profound changes in solutions only occur if the sales of a given crop provide a higher return than sugar beets at the favorable price. This happened in 2007, with wheat's price more than doubling in a 12-month period. At the moment of this writing, the current costs and prices are as follows (rounded figures):

	Wheat	Corn	Sugar Beets
Yield (T/acre)	2.5	3	20
Planting cost (\$/acre)	180	280	310
Selling price (\$/T)	300	170	41 under 6000 T 11 above 6000 T

The increase in wheat's selling price is due to a strong demand and low yields in Asia. These conditions may not prevail next year. Consider a model with a random selling price of wheat being 300 or 220 with equal probability. Purchase prices are as before 40% higher than selling prices. Compare the optimal solution with that of Table 5. How much would a farmer be willing to pay for a perfect forecast on the selling price of wheat?

9. If prices are also random variables, the news vendor's problem becomes more complicated. However, if prices and demands are independent random variables, show that the solution of the news vendor's problem is the one obtained before, where q and r are replaced by their expected values. Indicate under which conditions the same proposition is true for the farmer's problem.

10. In the news vendor's problem, we have assumed for simplicity that the random variable takes value from $-\infty$ to $+\infty$. Show that the optimal decisions are insensitive to this assumption, so that if the random variables have a nonzero density on a limited interval then the optimal solutions are obtained by the same analytical expression.

11. Suppose $c = 10$, $q = 25$, $r = 5$, and demand is uniform on $[50, 150]$. Find the optimal solution of the news vendor problem. Also, find the optimal solution of the deterministic model obtained by assuming a demand of 100. What is the value of the stochastic solution?

1.2 Financial Planning and Control

Financial decision-making problems can often be modeled as stochastic programs. In fact, the essence of financial planning is the incorporation of risk into investment decisions. The area represents one of the largest application areas of stochastic programming. Many references can be found in, for example, Mulvey and Vladimirou [1989, 1991b, 1992], Ziemba and Vickson [1975], and Zenios [1993].

We consider a simple example that illustrates additional stochastic programming properties. As in the farming example of Section 1.1, this example involves randomness in the constraint matrix instead of the right-hand side elements. These random variables reflect uncertain investment yields.

This section's example also has the characteristic that decisions are highly dependent on past outcomes. In the following capacity expansion problem of Section 1.3, this is not the case. In Chapter 3, we define this difference by a block separable recourse property that is present in some capacity expansion and similar problems.

For the current problem, suppose we wish to provide for a child's college education Y years from now. We currently have $\$b$ to invest in any of I investments. After Y years, we will have a wealth that we would like to have exceed a tuition goal of $\$G$. We suppose that we can change investments every υ years, so we have $H = Y/\upsilon$ investment periods. For our purposes here, we ignore transaction costs and taxes on income although these considerations would be important in reality. We also assume that all figures are in constant dollars.

In formulating the problem, we must first describe our objective in mathematical terms. We suppose that exceeding $\$ \ G$ after Y years would be equivalent to our having an income of $q\%$ of the excess while not meeting the goal would lead to borrowing for a cost $r\%$ of the amount short. This gives us the concave utility function in Figure 2. Many other forms of nonlinear utility functions are, of course, possible. See Kallberg and Ziemba [1983] for a description of their relevance in financial planning.

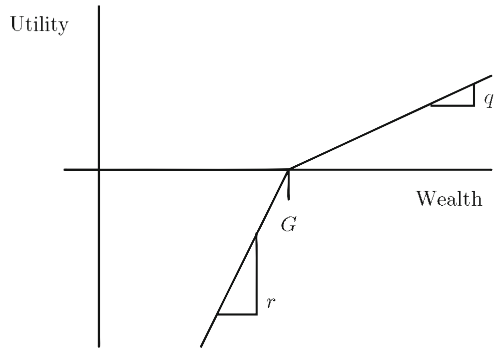

Fig. 2 Utility function of wealth at year Y for a goal G.

The major uncertainty in this model is the return on each investment i within each period t. We describe this random variable as $\boldsymbol{\xi}(i,t) = \xi(i,t,\omega)$ where ω is some underlying random element. The decisions on investments will also be random. We describe these decisions as $\mathbf{x}(i,t) = x(i,t,\omega)$. From the randomness of the returns and investment decisions, our final wealth will also be a random variable.

A key point about this investment model is that we cannot completely observe the random element ω when we make all our decisions $x(i,t,\omega)$. We can only observe the returns that have already taken place. In stochastic programming, we say that we cannot *anticipate* every possible outcome so our decisions are *nonanticipative* of future outcomes. Before the first period, this restriction corresponds to saying that we must make fixed investments, $x(i,1)$, for all $\omega \in \Omega$, the space of all random elements or, more specifically, returns that could possibly occur.

To illustrate the effects of including stochastic outcomes as well as modeling effects from choosing the time horizon Y and the coarseness of the period approximations H, we use a simple example with two possible investment types, stocks ($i = 1$) and government securities (bonds) ($i = 2$). We begin by setting Y at 15 years and allow investment changes every five years so that $H = 3$.

We assume that, over the three decision periods, eight possible scenarios may occur. The scenarios correspond to independent and equal likelihoods of having (inflation-adjusted) returns of 1.25 for stocks and 1.14 for bonds or 1.06 for stocks and 1.12 for bonds over the five-year period. We indicate the scenarios by an index $s = 1,\ldots,8$, which represents a collection of the outcomes ω that have common characteristics (such as returns) in a specific model. When we wish to allow more general interpretations of the outcomes, we use the base element ω. With the scenarios defined here, we assign probabilities for each s, $p(s) = 0.125$. The returns are $\xi(1,t,s) = 1.25$, $\xi(2,t,s) = 1.14$ for $t = 1, s = 1,\ldots,4$, for $t = 2$,

$s = 1,2,5,6$, and for $t = 3$, $s = 1,3,5,7$. In the other cases, $\xi(1,t,s) = 1.06$, $\xi(2,t,s) = 1.12$.

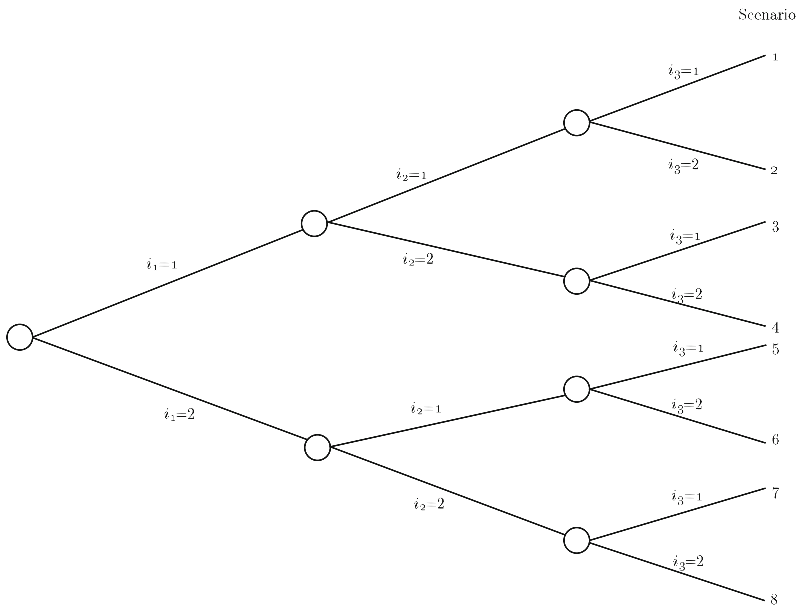

Fig. 3 Tree of scenarios for three periods.

The eight scenarios are represented by the tree in Figure 3. The scenario tree divides into branches corresponding to different realizations of the random returns. Because Scenarios 1 to 4, for example, have the same return for $t = 1$, they all follow the same first branch. Scenarios 1 and 2 then have the same second branch and finally divide completely in the last period. To show this more explicitly, we may refer to each scenario by the history of returns indexed by s_t for periods $t = 1,2,3$ as indicated on the tree in Figure 3. In this way, Scenario 1 may also be represented as $(s_1, s_2, s_3) = (1,1,1)$.

With the tree representation, we need only have a decision vector for each node of the tree. The decisions at $t = 1$ are just $x(1,1)$ and $x(2,1)$ for the amounts invested in stocks (1) and bonds (2) at the outset. For $t = 2$, we would have $x(i,2,s_1)$ where $i = 1,2$ for the type of investment and $s_1 = 1,2$ for the first-period return outcome. Similarly, the decisions at $t = 3$ are $x(i,3,s_1,s_2)$.

With these decision variables defined, we can formulate a mathematical program to maximize expected utility. Because the concave utility function in Figure 1 is piecewise linear, we just need to define deficit or shortage and excess or surplus variables, $w(i_1,i_2,i_3)$ and $y(i_1,i_2,i_3)$, and we can maintain a linear model. The objective is simply a probability- and penalty-weighted sum of these terms, which, in general, becomes:

$$\sum_{s_H} \cdots \sum_{s_1} p(s_1, \ldots, s_H)(-rw(s_1, \ldots, s_H) + qy(s_1, \ldots, s_H)) \ .$$

The first-period constraint is simply to invest the initial wealth:

$$\sum_i x(i,1) = b \ .$$

The constraints for periods $t = 2, \ldots, H$ are, for each $s_1, \ldots, s_{t-1}$:

$$\sum_i -\xi(i, t-1, s_1, \ldots, s_{t-1}) x(i, t-1, s_1, \ldots, s_{t-2}) + \sum_i x(i, t, s_1, \ldots, s_{t-1}) = 0 \ ,$$

while the constraints for period H are:

$$\sum_i \xi(i, H, s_1, \ldots, s_H) x(i, H, s_1, \ldots, s_{H-1}) - y(s_1, \ldots, s_H) + w(s_1, \ldots, s_H) = G \ .$$

Other constraints restrict the variables to be non-negative.

To specify the model in this example, we use initial wealth, $b = 55,000$; target value, $G = 80,000$; surplus reward, $q = 1$; and shortage penalty, $r = 4$. The result is a stochastic program in the following form where the units are thousands of dollars:

$$\max z = \sum_{s_1=1}^{2} \sum_{s_2=1}^{2} \sum_{s_3=1}^{2} 0.125(y(s_1, s_2, s_3) - 4w(s_1, s_2, s_3)) \tag{2.1}$$

$$\begin{aligned}
\text{s. t.} \quad & x(1,1) + x(2,1) &= 55 \ , \\
& -1.25x(1,1) - 1.14x(2,1) + x(1,2,1) + x(2,2,1) &= 0 \ , \\
& -1.06x(1,1) - 1.12x(2,1) + x(1,2,2) + x(2,2,2) &= 0 \ , \\
& -1.25x(1,2,1) - 1.14x(2,2,1) + x(1,3,1,1) + x(2,3,1,1) &= 0 \ , \\
& -1.06x(1,2,1) - 1.12x(2,2,1) + x(1,3,1,2) + x(2,3,1,2) &= 0 \ , \\
& -1.25x(1,2,2) - 1.14x(2,2,2) + x(1,3,2,1) + x(2,3,2,1) &= 0 \ , \\
& -1.06x(1,2,2) - 1.12x(2,2,2) + x(1,3,2,2) + x(2,3,2,2) &= 0 \ , \\
& 1.25x(1,3,1,1) + 1.14x(2,3,1,1) - y(1,1,1) + w(1,1,1) &= 80 \ , \\
& 1.06x(1,3,1,1) + 1.12x(2,3,1,1) - y(1,1,2) + w(1,1,2) &= 80 \ , \\
& 1.25x(1,3,1,2) + 1.14x(2,3,1,2) - y(1,2,1) + w(1,2,1) &= 80 \ , \\
& 1.06x(1,3,1,2) + 1.12x(2,3,1,2) - y(1,2,2) + w(1,2,2) &= 80 \ , \\
& 1.25x(1,3,2,1) + 1.14x(2,3,2,1) - y(2,1,1) + w(2,1,1) &= 80 \ , \\
& 1.06x(1,3,2,1) + 1.12x(2,3,2,1) - y(2,1,2) + w(2,1,2) &= 80 \ , \\
& 1.25x(1,3,2,2) + 1.14x(2,3,2,2) - y(2,2,1) + w(2,2,1) &= 80 \ , \\
& 1.06x(1,3,2,2) + 1.12x(2,3,2,2) - y(2,2,2) + w(2,2,2) &= 80 \ , \\
& x(i,t,s_1,\ldots,s_{t-1}) \geq 0 \ , \ y(s_1,s_2,s_3) \geq 0 \ , \ w(s_1,s_2,s_3) &\geq 0 \ , \\
& \text{for all } i,t,s_1,s_2,s_3 \ .
\end{aligned}$$

Solving the problem in (2.1) yields an optimal expected utility value of -1.514 . We call this value, RP , for the expected *recourse problem* solution value. The optimal solution (in thousands of dollars) appears in Table 6.

Table 6 Optimal solution with three-period stochastic program.

Period, Scenario	Stock	Bonds
1,1-8	41.5	13.5
2,1-4	65.1	2.17
2,5-8	36.7	22.4
3,1-2	83.8	0.00
3,3-4	0.00	71.4
3,5-6	0.00	71.4
3,7-8	64.0	0.00
Scenario	Above G	Below G
1	24.8	0.00
2	8.87	0.00
3	1.43	0.00
4	0.00	0.00
5	1.43	0.00
6	0.00	0.00
7	0.00	0.00
8	0.00	12.2

In this solution, the initial investment is heavily in stock ($41,500) with only $13,500 in bonds. Notice the reaction to first-period outcomes, however. In the case of Scenarios 1 to 4, stocks are even more prominent, while Scenarios 5 to 8 reflect a more conservative government security portfolio. In the last period, notice how the investments are either completely in stocks or completely in bonds. This is a general trait of one-period decisions. It occurs here because in Scenarios 1 and 2, there is no risk of missing the target. In Scenarios 3 to 6, stock investments may cause one to miss the target, so they are avoided. In Scenarios 7 and 8, the only hope of reaching the target is through stocks.

We compare the results in Table 6 to a deterministic model in which all random returns are replaced by their expectation. For that model, because the expected return on stock is 1.155 in each period, while the expected return on bonds is only 1.13 in each period, the optimal investment plan places all funds in stocks in each period. If we implement this policy each period, but instead observed the random returns, we would have an expected utility called the *expected value* solution, or EV . In this case, we would realize an expected utility of $EV = -3.788$, while the stochastic program value is again $RP = -1.514$. The difference between these quantities is the value of the stochastic solution:

$$VSS = RP - EV = -1.514 - (-3.788) = 2.274 \,.$$

This comparison gives us a measure of the utility value in using a decision from a stochastic program compared to a decision from a deterministic program. Another comparison of models is in terms of the probability of reaching the goal. Models with these types of objectives are called *chance-constrained programs* or *programs with probabilistic constraints* (see Charnes and Cooper [1959] and Prékopa [1973]). Notice that the stochastic program solution reaches the goal 87.5% of the time. The expected value deterministic model solution only reaches the goal 50% of the time. In this case, the value of the stochastic solution may be even more significant.

The formulation we gave in (2.1) can become quite cumbersome as the time horizon, H, increases and the decision tree of Figure 3 grows quite bushy. Another modeling approach to this type of multistage problem is to consider the full horizon scenarios, s, directly, without specifying the history of the process. We then substitute a scenario set S for the random elements Ω. Probabilities, $p(s)$, returns, $\xi(i,t,s)$, and investments, $x(i,t,s)$, become functions of the H-period scenarios and not just the history until period t.

The difficulty is that, when we have split up the scenarios, we may have lost nonanticipativity of the decisions because they would now include knowledge of the outcomes up to the end of the horizon. To enforce nonanticipativity, we add constraints explicitly in the formulation. First, the scenarios that correspond to the same set of past outcomes at each period form groups, $S^t_{s_1,\dots,s_{t-1}}$, for scenarios at time t. Now, all actions up to time t must be the same within a group. We do this through an explicit constraint. The new general formulation of (2.1) becomes:

$$
\begin{aligned}
\max z = \sum_s p(s)(qy(s) - rw(s)) & \\
\text{s. t. } \sum_{i=1}^{I} x(i,1,s) = b\ ,\ \forall s \in S\ , & \qquad (2.2)\\
\sum_{i=1}^{I} \xi(i,t,s)x(i,t-1,s) - \sum_{i=1}^{I} x(i,t,s) = 0\ ,\ \forall s \in S\ , & \\
t = 2,\dots,H\ , & \\
\sum_{i=1}^{I} \xi(i,H,s)x(i,H,s) - y(s) + w(s) = G\ , & \\
\left(\sum_{s' \in S^t_{J(s,t)}} p(s')x(i,t,s')\right) - \left(\sum_{s' \in S^t_{J(s,t)}} p(s')\right) x(i,t,s) = 0\ , & \\
\forall 1 \le i \le I\ ,\ \forall 1 \le t \le H\ ,\ \forall s \in S\ , & \\
x(i,t,s) \ge 0\ ,\ y(s) \ge 0\ ,\ w(s) \ge 0\ , & \\
\forall\ 1 \le i \le I\ ,\ \forall\ 1 \le t \le H\ ,\ \forall\ s \in S\ , &
\end{aligned}
$$

where $J(s,t) = \{s_1,\dots,s_{t-1}\}$ such that $s \in S^t_{s_1,\dots,s_{t-1}}$. Note that the last equality constraint indeed forces all decisions within the same group at time t to be the same. Formulation (2.2) has a special advantage for the problem here because these *nonanticipativity* constraints are the only constraints linking the separate scenarios. Without them, the problem would decompose into a separate problem for each s, maintaining the structure of that problem.

In modeling terms, this simple additional constraint makes it relatively easy to move from a deterministic model to a stochastic model of the same problem. This ease of conversion can be especially useful in modeling languages. For example, Figure 4 gives a complete AMPL (Fourer, Gay, and Kernighan [1993]) model of the problem in (2.2). In this language, *set*, *param*, and *var* are keywords for sets, parameters, and variables. The addition of the scenario indicators and nonanticipativity constraints (*nonanticip*) are the only additions to a deterministic model.

```
# This problem describes a simple financial planning problem
# for financing college education
set investments; # different investment options
param initwealth; # initial holdings
param H; # number of periods
param scenarios; # number of scenarios (total S)
# The following 0-1 array shows which scenarios are combined at period H
param scen_links { 1..scenarios,1..scenarios,1..H } ;
param target; # target value G at time H
param invest; # value of investing beyond target value
param penalty; # penalty for not meeting target
param return { investments,1..scenarios,1..H } ; # return on each inv
param prob { 1..scenarios } ; # probability of each scenario
# variables
var amtinvest { investments,1..scenarios,1..H } ¿= 0; #actual amounts inv'd
var above_target { 1..scenarios } ¿= 0; # amt above final target
var below_target { 1..scenarios } ¿= 0; # amt below final target
# objective
maximize exp_value : sum { i in 1..scenarios } prob[i]*(invest*above_target[i]
- penalty*below_target[i]);
# constraints
subject to budget { i in 1..scenarios } :
sum { k in investments } (amtinvest[k,i,1]) = initwealth;#invest initial wealth
subject to nonanticip { k in investments,j in 1..scenarios,t in 1..H } :
(sum { i in 1..scenarios } scen_links[j,i,t]*prob[i]*amtinvest[k,i,t]) -
(sum { i in 1..scenarios } scen_links[j,i,t]*prob[i])*
amtinvest[k,j,H] = 0; # makes all investments nonanticipative
subject to balance { j in 1..scenarios, t in 1..H-1 } :
(sum { k in investments } return[k,j,t]*amtinvest[k,j,t]) - sum { k in
investments } amtinvest[k,j,t+1] = 0; # reinvest each time period
subject to scenario_value { j in 1..scenarios } : (sum { k in
investments } return[k,j,H]*amtinvest[k,j,H]) - above_target[j] +
below_target[j] = target; # amounts not meeting target
```

Fig. 4 AMPL format of financial planning model.

Given the ease of this modeling effort, standard optimization procedures can be simply applied to this problem. However, as we noted earlier, the number of scenarios can become extremely large. Standard methods may not be able to solve the problem in any reasonable amount of time, necessitating other techniques. The remaining chapters in this book focus on these other methods and on procedures for creating models that are amenable to those specialized techniques.

In financial problems, it is particularly worthwhile to try to exploit the underlying structure of the problem without the nonanticipativity constraints. This relaxed problem is in fact a *generalized network* that allows the use of efficient network optimization methods that cannot apply to the full problem in (2.2). We discuss this option more thoroughly in Chapter 5.

With either formulation (2.1) or (2.2), in completing the model, some decisions must be made about the possible set of outcomes or scenarios and the coarseness of the period structure, i.e., the number of periods H allowed for investments. We must also find probabilities to attach to outcomes within each of these periods. These probabilities are often approximations that can, as we shall see in Chapter 8, provide bounds on true values or on uncertain outcomes with incompletely known distributions. A key observation is that the important step is to include stochastic elements at least approximately and that deterministic solutions most often give misleading results.

In closing this section, note that the mathematical form of this problem actually represents a broad class of control problems (see, for example, Varaiya and Wets [1989]). In fact, it is basically equivalent to any control problem governed by a linear system of differential equations. We have merely taken a discrete time approach to this problem. This approach can be applied to the control of a wide variety of electrical, mechanical, chemical, and economic systems. We merely redefine state variables (now, wealth) in each time period and controls (investment levels). The random gain or loss is reflected in the return coefficients. Typically, these types of control problems would have nonlinear (e.g., *quadratic*) costs associated with the control in each time period. This presents no complication for our purposes, so we may include any of these problems as potential applications. In Section 1.4, we will look at a fundamentally nonlinear problem in more detail.

Exercises

1. Suppose you consider just a five-year planning horizon. Choose an appropriate target and solve over this horizon with a single first-period decision.
2. Suppose you implement a buy-and-hold strategy and make a single investment decision without any additional trading until the end of the time horizon. Formulate and solve this problem to determine an optimal allocation.
3. Suppose that goal G is also a random parameter and could be \$75,000 or \$85,000 with equal probabilities. Formulate and solve this problem. Compare this solution to the solution for the problem with a known target.
4. Suppose that every trade (purchase or sale) of an asset involves a transaction cost that is equal to 1% of the amount traded. Re-formulate the problem with this transaction cost and solve for the optimal solution.

1.3 Capacity Expansion

Capacity expansion models optimal choices of the timing and levels of investments to meet future demands of a given product. This problem has many applications. Here we illustrate the case of power plant expansion for electricity generation: we want to find optimal levels of investment in various types of power plants to meet future electricity demand.

We first present a *static deterministic analysis* of the electricity generation problem. *Static* means that decisions are taken only once. *Deterministic* means that the future is supposed to be fully and perfectly known.

Three properties of a given power plant i can be singled out in a static analysis: the investment cost r_i, the operating cost q_i, and the availability factor a_i, which indicates the percent of time the power plant can effectively be operated. Demand for electricity can be considered a single product, but the level of demand varies over time. Analysts usually represent the demand in terms of a so-called *load duration curve* that describes the demand over time in decreasing order of demand level (Figure 5). The curve gives the time, τ, that each demand level, D, is reached. Because here we are concerned with investments over the long run, the load duration curve we consider is taken over the life cycle of the plants.

The load duration curve can be approximated by a piecewise constant curve (Figure 6) with m segments. Let $d_1 = D_1$, $d_j = D_j - D_{j-1}$, $j = 2, \ldots, m$ represent the additional power demanded in the so-called *mode j* for a duration τ_j. To obtain a good approximation of the load curve, it is necessary to consider large values of m. In the static situation, the problem consists of finding the optimal investment for each mode j, i.e., to find the particular type of power plant i, $i = 1, \ldots, n$, that minimizes the total cost of effectively producing 1 MW (megawatt) of electricity during the time τ_j. It is given by

$$i(j) = \text{argmin}_{i=1,\ldots,n} \left\{ \frac{r_i + q_i \; \tau_j}{a_i} \right\} , \tag{3.1}$$

where n is the number of available technologies and argmin represents the index i for which the minimum is achieved.

The static model (3.1) captures one essential feature of the problem, namely, that base load demand (associated with large values of τ_j, i.e., small indices j) is covered by equipment with low operating costs (scaled by availability factor), while peak-load demand (associated with small values of τ_j, i.e., large indices j) is covered by equipment with low investment costs (also scaled by their availability factor). For the sake of completeness, peak-load equipment should also offer operational flexibility.

At least four elements justify considering a *dynamic* or *multistage model* for the electricity generation investment problem:

- the long-term evolution of equipment costs;
- the long-term evolution of the load curve;

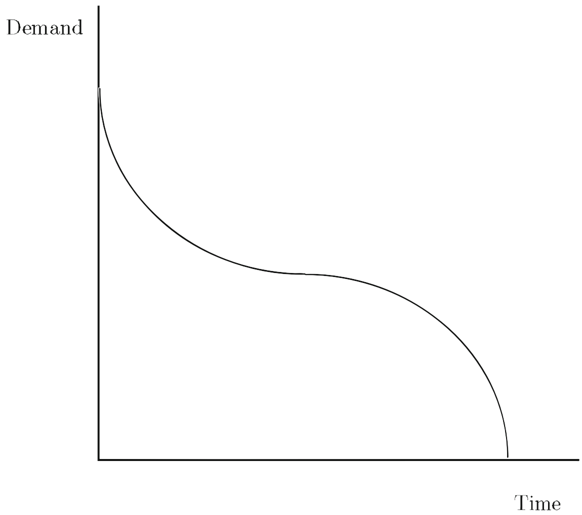

Fig. 5 The load duration curve.

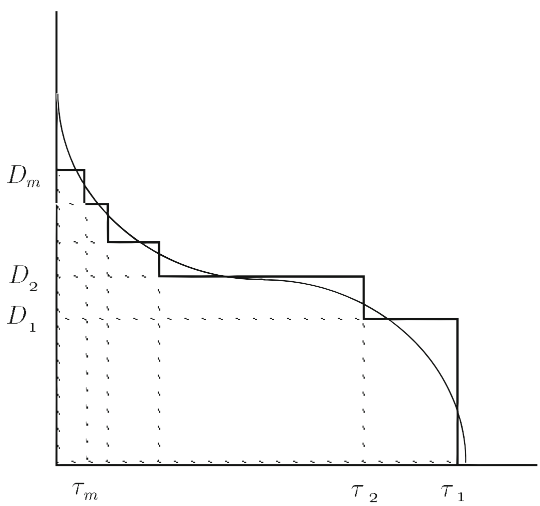

Fig. 6 A piecewise constant approximation of the load duration curve.

- the appearance of new technologies;
- the obsolescence of currently available equipment.

The equipment costs are influenced by technological progress but also (and, for some, drastically) by the evolution of fuel costs.

Of significant importance in the evolution of demand is both the total energy demanded (the area under the load curve) and the peak-level D_m , which determines the total capacity that should be available to cover demand. The evolution of the load curve is determined by several factors, including the level of activity in industry, energy savings in general, and the electricity producers' rate policy.

The appearance of new technologies depends on the technical and commercial success of research and development while obsolescence of available equipment depends on past decisions and the technical lifetime of equipment. All the elements together imply that it is no longer optimal to invest only in view of the short-term ordering of equipment given by (3.1) but that a long-term optimal policy should be found.

The following multistage model can be proposed. Let

- $t = 1,\ldots,H$ index the periods or stages;
- $i = 1,\ldots,n$ index the available technologies;
- $j = 1,\ldots,m$ index the operating modes in the load duration curve.

Also define the following:

- $a_i =$ availability factor of i ;
- $L_i =$ lifetime of i ;
- $g_i^t =$ existing capacity of i at time t , decided before $t = 1$;
- $r_i^t =$ unit investment cost for i at time t (assuming a fixed plant life cycle for each type i of plant);
- $q_i^t =$ unit production cost for i at time t ;
- $d_j^t =$ maximal power demanded in mode j at time t ;
- $\tau_j^t =$ duration of mode j at time t .

Consider, finally, the set of decisions

- $x_i^t =$ new capacity made available for technology i at time t ;
- $w_i^t =$ total capacity of i available at time t ;
- $y_{ij}^t =$ capacity of i effectively used at time t in mode j .

The electricity generation H-stage problem can be defined as

$$\min_{x,y,w} \sum_{t=1}^{H} \left(\sum_{i=1}^{n} r_i^t \cdot w_i^t + \sum_{i=1}^{n} \sum_{j=1}^{m} q_i^t \cdot \tau_j^t \cdot y_{ij}^t \right) \tag{3.2}$$

$$\text{s. t. } w_i^t = w_i^{t-1} + x_i^t - x_i^{t-L_i} , \qquad i = 1,\ldots,n , \ t = 1,\ldots,H , \tag{3.3}$$

$$\sum_{i=1}^{n} y_{ij}^t = d_j^t , \qquad j = 1,\ldots,m , \ t = 1,\ldots,H , \tag{3.4}$$

$$\sum_{j=1}^{m} y_{ij}^t \leq a_i(g_i^t + w_i^t) , \qquad i = 1,\ldots,n , \ t = 1,\ldots,H , \tag{3.5}$$

$$x, y, w \geq 0 .$$

Decisions in each period t involve new capacities x_i^t made available in each technology and capacities y_{ij}^t operated in each mode for each technology.

Newly decided capacities increase the total capacity w_i^t made available, as given by (3.3), where the equipment's becoming obsolete after its lifetime is also considered. We assume $x_i^t = 0$ if $t \leq 0$, so equation (3.3) only involves newly decided capacities.

By (3.4), the optimal operation of equipment must be chosen to meet demand in all modes using available capacities, which by (3.5) depend on capacities g_i^t decided before $t = 1$, newly decided capacities x_i^t, and the availability factor.

The objective function (3.2) is the sum of the investment plus maintenance costs and operating costs. Compared to (3.1), availability factors enter constraints (3.5) and do not need to appear in the objective function. The operating costs are exactly the same and are based on operating decisions y_{ij}^t, while the investment annuities and maintenance costs r_i^t apply on the cumulative capacity w_i^t. Placing annuities on the cumulative capacity, instead of charging the full investment cost to the decision x_i^t, simplifies the treatment of end of horizon effects and is currently used in many power generation models. It is a special case of the salvage value approach and other period aggregations discussed in Section 10.2.

The same reasons that plead for the use of a multistage model motivate resorting to a *stochastic model*. The evolution of equipment costs, particularly fuel costs, the evolution of total demand, the date of appearance of new technologies, even the lifetime of existing equipment, can all be considered truly random. The main difference between the stochastic model and its deterministic counterpart is in the definition of the variables x_i^t and w_i^t. In particular, x_i^t now represents the new capacity of i decided at time t, which becomes available at time $x_i^{t+\Delta_i}$, where Δ_i is the construction delay for equipment i. In other words, to have extra capacity available at time t, it is necessary to decide at $t - \Delta_i$, when less information is available on the evolution of demand and equipment costs. This is especially important because it would be preferable to be able to wait until the last moment to take decisions that would have immediate impact.

Assume that each decision is now a random variable. Instead of writing an explicit dependence on the random element, ω, we again use boldface notation to denote random variables. We then have:

- $\mathbf{x}_i^t =$ new capacity decided at time t for equipment i, $i = 1,\ldots,n$;
- $\mathbf{w}_i^t =$ total capacity of i available and in order at time t;
- $\boldsymbol{\xi} =$ the vector of random parameters at time t;

and all other variables as before. The stochastic model is then

$$\min \mathrm{E}_{\boldsymbol{\xi}} \sum_{t=1}^{H} \left(\sum_{i=1}^{n} \mathbf{r}_i^t \mathbf{w}_i^t + \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbf{q}_i^t \, \tau_j^t \, \mathbf{y}_{ij}^t \right) \tag{3.6}$$

$$\text{s. t. } \mathbf{w}_i^t = \mathbf{w}_i^{t-1} + \mathbf{x}_i^t - \mathbf{x}_i^{t-L_i} \,, \qquad i=1,\dots,n \,,\ t=1,\dots,H \,, \tag{3.7}$$

$$\sum_{i=1}^{n} \mathbf{y}_{ij}^t = \mathbf{d}_j^t \,, \qquad j=1,\dots,m \,,\ t=1,\dots,H \,, \tag{3.8}$$

$$\sum_{j=1}^{m} \mathbf{y}_{ij}^t \le a_i (g_i^t + \mathbf{w}_i^{t-\Delta_i}) \,, \qquad i=1,\dots,n \,,\ t=1,\dots,H \,, \tag{3.9}$$

$$\mathbf{w}, \mathbf{x}, \mathbf{y} \ge 0 \,,$$

where the expectation is taken with respect to the random vector $\boldsymbol{\xi} = (\boldsymbol{\xi}^2, \dots, \boldsymbol{\xi}^H)$. Here, the elements forming $\boldsymbol{\xi}^t$ are the demands, $(\mathbf{d}_1^t, \dots, \mathbf{d}_k^t)$, and the cost vectors, $(\mathbf{r}^t, \mathbf{q}^t)$. In some cases, $\boldsymbol{\xi}^t$ can also contain the lifetimes L_i, the delay factors Δ_i, and the availability factors a_i, depending on the elements deemed uncertain in the future.

Formulation (3.6)–(3.9) is a convenient representation of the stochastic program. At some point, however, this representation might seem a little confusing. For example, it seems that the expectation is taken only on the objective function, while the constraints contain random coefficients (such as $\mathbf{d}_j^t$ in the right-hand side of (3.8)).

Another important aspect is the fact that decisions taken at time t, $(\mathbf{w}^t, \mathbf{y}^t)$, are dependent on the particular realization of the random vector, $\boldsymbol{\xi}^t$, but cannot depend on future realizations of the random vector. This is clearly a desirable feature for a truly stochastic decision process. If demands in several periods are high, one would expect investors to increase capacity much more than if, for example, demands remain low.

Formally, if the decision variables $(\mathbf{w}^t, \mathbf{y}^t)$ were not dependent on $\boldsymbol{\xi}^t$, the objective function in (3.6) could be replaced by

$$\sum_t \sum_i \left(\mathrm{E}_{\boldsymbol{\xi}} \, \mathbf{r}_i^t \, w_i^t + \sum_j \mathrm{E}_{\boldsymbol{\xi}} \, \mathbf{q}_i^t \, \tau_i^t \, y_{ij}^t \right) = \sum_t \sum_i \left(\bar{r}_i^t \cdot w_i^t + \sum_j (\overline{q_i \tau_j}) y_{ij}^t \right) \,, \tag{3.10}$$

where $\bar{r}_i^t = \mathrm{E}_{\boldsymbol{\xi}} \mathbf{r}_i^t$ and $\overline{q_i \tau_j} = \mathrm{E}_{\boldsymbol{\xi}} (\mathbf{q}_i^t \, \tau_j^t)$, making problem (3.6) to (3.9) deterministic. In the next section, we will make the dependence of the decision variables on the random vector explicit.

The formulation given earlier is convenient in its allowing for both continuous and discrete random variables. Theoretical properties such as continuity and convexity can be derived for both types of variables. Solution procedures, on the other hand, strongly differ.

Problem (3.6) to (3.9) is a multistage stochastic linear program with several random variables that actually has an additional property, called *block separable recourse*. This property stems from a separation that can be made between the aggregate-level decisions, $(\mathbf{x}^t, \mathbf{w}^t)$, and the detailed-level decisions, $\mathbf{y}^t$.

We will formally define block separability in Chapter 3, but we can make an observation about its effect here. Suppose future demands are always independent of the past. In this case, the decision on capacity to install in the future at some t only depends on available capacity and does not depend on the outcomes up to time t. The same $\mathbf{x}^t$ must then be optimal for any realization of $\boldsymbol{\xi}$. The only remaining stochastic decision is in the operation-level vector, $\mathbf{y}^t$, which now depends separately on each period's capacity. The overall result is that a multiperiod problem now becomes a much less complex two-period problem.

As a simple example, consider the following problem that appears in Louveaux and Smeers [1988]. In this case, the resulting two period model has three operating modes, $n = 4$ technologies, $\Delta_i = 1$ period of construction delay, full availabilities, $a \equiv 1$, and no existing equipment, $g \equiv 0$. The only random variable is $\mathbf{d}_1 = \boldsymbol{\xi}$. The other demands are $d_2 = 3$ and $d_3 = 2$. The investment costs are $r^1 = (10,7,16,6)^T$ with production costs $q^2 = (4,4.5,3.2,5.5)^T$ and load durations $\tau^2 = (10,6,1)^T$. We also add a budget constraint to keep all investment below 120. The resulting two-period stochastic program is:

$$
\begin{aligned}
\min\ & 10x_1^1 + 7x_2^1 + 16x_3^1 + 6x_4^1 + \mathrm{E}_{\boldsymbol{\xi}}\Big[\sum_{j=1}^{3} \tau_j^2 (4\mathbf{y}_{1j}^2 + 4.5\mathbf{y}_{2j}^2 \\
& \qquad\qquad\qquad\qquad\qquad\qquad + 3.2\mathbf{y}_{3j}^2 + 5.5\mathbf{y}_{4j}^2)\Big] \\
\text{s. t. }\ & 10x_1^1 + 7x_2^1 + 16x_3^1 + 6x_4^1 \le 120\,, \qquad\qquad (3.11)\\
& -x_i^1 + \sum_{j=1}^{3} \mathbf{y}_{ij}^2 \le 0\,, \qquad i = 1,\ldots,4\,,\\
& \sum_{i=1}^{y} \mathbf{y}_{i1}^2 = \boldsymbol{\xi}\,,\\
& \sum_{i=1}^{y} \mathbf{y}_{ij}^2 = d_j^2\,, \qquad j = 2,3\,,\\
& x_1^1 \ge 0\,, \quad x_2^1 \ge 0\,, \quad x_3^1 \ge 0\,, \quad x_4^1 \ge 0\,,\\
& \mathbf{y}_{ij}^2 \ge 0\,, \quad i = 1,\ldots,4\,, \quad j = 1,2,3\,.
\end{aligned}
$$

Assuming that $\boldsymbol{\xi}$ takes on the values 3, 5, and 7 with probabilities 0.3, 0.4, and 0.3, respectively, an optimal stochastic programming solution to (3.11) includes $x^{1*} = (2.67,4.00,3.33,2.00)^T$ with an optimal objective value of 381.85. We can again consider the expected value solution, which would substitute $\boldsymbol{\xi} \equiv 5$ in (3.11). An optimal solution here (again not unique) is $\bar{x}^1 = (0.00,3.00,5.00,2.00)^T$. The objective value, if this single event occurs, is 365. However, if we use this solution in the stochastic problem, then with probability 0.3, demand cannot be met. This would yield an infinite value of the stochastic solution.

Infinite values probably do not make sense in practice because an action can be taken somehow to avoid total system collapse. The power company could buy from neighboring utilities, for example, but the cost would be much higher than any company operating cost. An alternative technology (internal or external to the company) that is always available at high cost is called a *backstop* technology. If we assume, for example, in problem (3.11) that some other technology is always available, without any required investment costs at a unit operating cost of 100, then the expected value solution would be feasible and have an expected stochastic program value of 427.82. In this case, the value of the stochastic solution becomes $427.82 - 381.85 = 45.97$.

In many power problems, focus is on the reliability of the system or the system's ability to meet demand. This reliability is often described as expressing a minimum probability for meeting demand using the non-backstop technologies. If these technologies are $1, \ldots, n-1$, then the reliability restriction (in the two-period situation where capacity decisions need not be random) is:

$$\mathrm{P}[\sum_{i=1}^{n-1} a_i(g_i^t + w_i^t) \geq \sum_{j=1}^{m} \mathbf{d}_j^t] \geq \alpha \;, \qquad \forall t \;, \tag{3.12}$$

where $0 < \alpha \leq 1$. Inequality (3.12) is called a *chance* or *probabilistic constraint* in stochastic programming. In production problems, these constraints are often called *fill rate* or *service rate constraints*. They place restrictions on decisions so that constraint violations are not too frequent. Hence, we would often have α quite close to 1.

If the only probabilistic constraints are of the form in (3.12), then we simply want the cumulative available capacity at time t to be at least the α quantile of the cumulative demand in all modes at time t. We then obtain a *deterministic equivalent* constraint to (3.12) of the following form:

$$\sum_{i=1}^{n-1} a_i(g_i^t + w_i^t) \geq (F^t)^{-1}(\alpha) \;, \qquad \forall t \;, \tag{3.13}$$

where F^t is the (assumed continuous) distribution function of $\sum_{j=1}^{m} \mathbf{d}_j^t$ and $F^{-1}(\alpha)$ is the α-quantile of F. Constraints of the form in (3.13) can then be added to (3.6) to (3.9) or, indeed, to the deterministic problem in (3.2) to (3.5), where expected values replace the random variables.

By adding these chance constraint equivalents, many of the problems of deterministic formulations can be avoided. For example, if we choose $\alpha = 0.7$ for the problem in (3.11), then adding a constraint of the form in (3.13) would not change the deterministic expected value solution. However, we would get a different result if we set $\alpha = 1.0$. In this case, constraint (3.13) for the given data becomes simply:

$$\sum_{i=1}^{4} w_i^1 \geq 12 \;. \tag{3.14}$$

Adding (3.14) to the expected value problem results in an optimal solution with $w^{1*} = (0.833, 3.00, 4.17, 4.00)^T$. The expected value of using this solution in the stochastic program is 383.99, or only 2.14 more than the optimal value in (3.11).

In general, probabilistic constraints are represented by deterministic equivalents and are often included in stochastic programs. We discuss some of the theory of these constraints in Chapter 3. Our emphasis in this book is, however, on optimizing the expected value of continuous utility functions, such as the costs in this capacity expansion problem. We, therefore, concentrate on recourse problems and assume that probabilistic constraints are represented by deterministic equivalents within our formulations.

This problem illustrates a multistage decision problem and the addition of probabilistic constraints. The structure of the problem, however, allows for a two-stage equivalent problem. In this way, the capacity expansion problem provides a bridge between the two-stage example of Section 1.1 and the multistage problem of Section 1.2.

This problem also has a natural interpretation with discrete decision variables. For most producing units, only a limited number of possible sizes exists. Typical sizes for high-temperature nuclear reactors would be 1000 MW and 1300 MW, so that capacity decisions could only be taken as integer multiples of these values.

Exercises

1. The detailed-level decisions can be found quite easily according to an *order of merit rule*. In this case, one begins with Mode 1 and uses the least expensive equipment until its capacity is exhausted or demand is satisfied. One continues to exhaust capacity or satisfy demand in order of increasing unit operating cost and mode. Show that this procedure is indeed optimal for determining the y^t_{ij} values.
2. Prove that, in the case of no serial correlation ($\boldsymbol{\xi}^t$ and $\boldsymbol{\xi}^{t+1}$ stochastically independent), an optimal solution has the same value for $\mathbf{w}^t$ and $\mathbf{x}^t$ for all $\boldsymbol{\xi}$. Give an example where this does not occur with serial correlation.
3. For the example in (3.11), suppose we add a reliability constraint of the form in (3.14) to the expected value problem, but we use a right-hand side of 11 instead of 12. What is the stochastic program expected value of this solution?

1.4 Design for Manufacturing Quality

This section illustrates a common engineering problem that we model as a stochastic program. The problem demonstrates nonlinear functions in stochastic programming and provides further evidence of the importance of the stochastic solution.

Consider a designer deciding various product specifications to achieve some measure of product cost and performance. The specifications may not, however, completely determine the characteristics of each manufactured product. Key characteristics of the product are often random. For example, every item includes variations due to machining or other processing. Each consumer also does not use the product in the same way. Cost and performance characteristics thus become random variables.

Deterministic methods may yield costly results that are only discovered after production has begun. From this experience, designing for quality and consideration of variable outcomes has become an increasingly important aspect of modern manufacturing (see, for example, Taguchi et al. [1989]). In industry, the methods of Taguchi have been widely used (see also Taguchi [1986]). Taguchi methods can, in fact, be seen as examples of stochastic programming, although they are often not described this way.

In this section, we wish to give a small example of the uses of stochastic programming in manufacturing design and to show how the general stochastic programming approach can be applied. We note that we base our analysis on actual performance measures, whereas the Taguchi methods generally attach surrogate costs to deviations from nominal parameter values.

We consider the design of a simple axle assembly for a bicycle cart. The axle has the general appearance in Figure 7.

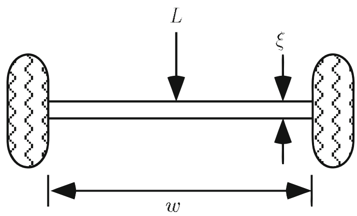

Fig. 7 An axle of length w and diameter ξ with a central load L.

The designer must determine the specified length w and design diameter ξ of the axle. We use inches to measure these quantities and assume that other dimensions are fixed. Together, these quantities determine the performance characteristics of the product. The goal is to determine a combination that gives the greatest expected profit.

The initial costs are for manufacturing the components. We assume that a single process is used for the two components. No alternative technologies are available, although, in practice, several processes might be available. When the axle is produced, the actual dimensions are not exactly those that are specified. For this example, we suppose that the length w can be produced exactly but that the diameter ξ is a random variable, $\boldsymbol{\xi}(x)$, that depends on a specified mean value, x, that represents, for example, the setting on a machine. We assume a triangular distribution for $\boldsymbol{\xi}(x)$ on $[0.9x, 1.1x]$. This distribution has a density,

$$f_x(\xi) = \begin{cases} (100/x^2)(\xi - 0.9x) & \text{if } 0.9x \le \xi < x\,, \\ (100/x^2)(1.1x - \xi) & \text{if } x \le \xi \le 1.1x\,, \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

The decision is then to determine w and x, subject to certain limits, $w \le w^{\max}$ and $x \le x^{\max}$, in order to maximize expected profits. For revenues, we assume that if the product is profitable, we sell as many as we can produce. This amount is fixed by labor and equipment regardless of the size of the axle. We, therefore, only wish to determine the maximum selling price that generates enough demand for all production. From marketing studies, we determine that this maximum selling price depends on the length and is expressed as

$$r(1 - e^{-0.1w})\,, \tag{4.2}$$

where r is the maximum possible for any such product.

Our production costs for labor and equipment are assumed fixed, so only material cost is variable. This cost is proportional to the mean values of the specified dimensions because material is acquired before the actual machining process. Suppose c is the cost of a single axle material unit. The total manufacturing cost for an item is then

$$c\left(\frac{w\pi x^2}{4}\right)\,. \tag{4.3}$$

In this simplified model, we assume that no quantity discounts apply in the production process.

Other costs are incurred after the product is made due to warranty claims and potential future sales losses from product defects. These costs are often called *quality losses*. In stochastic programming terms, these are the recourse costs. Here, the product may perform poorly if the axle becomes bent or broken due to excess stress or deflection. The stress limit, assuming a steel axle and 100-pound maximum central load, is

$$\frac{w}{\xi^3} \le 39.27\,. \tag{4.4}$$

For deflection, we use a maximum 2000-rpm speed (equivalent to a speed of 60 km/hour for a typical 15-centimeter wheel) to obtain:

$$\frac{w^3}{\xi^4} \le 63,169\,. \tag{4.5}$$

When either of these constraints is violated, the axle deforms. The expected cost for not meeting these constraints is assumed proportional to the square of the violation. We express it as

$$Q(w,x,\xi) = \min_y \{qy^2 \text{ s. t. } \frac{w}{\xi^3} - y \le 39.27, \frac{w^3}{\xi^4} - 300y \le 63,169\}\,, \tag{4.6}$$

where y is, therefore, the maximum of stress violation and (to maintain similar units) $\frac{1}{300}$ of the deflection violation.

The expected cost, given w and x, is

$$\mathscr{Q}(w,x) = \int_{\xi} Q(w,x,\xi) f_x(\xi) d\xi \ , \tag{4.7}$$

which can be written as:

$$\mathscr{Q}(w,x) = q \int_{.9x}^{1.1x} (100/x^2) \min\{\xi - .9x, 1.1x - \xi\} \\ [\max\{0, \left(\frac{w}{\xi^3}\right) - 39.27, \left(\frac{w^3}{300\xi^4}\right) - 210.56\}]^2 \mathrm{d}\xi \ . \tag{4.8}$$

The overall problem is to find:

$$\max \ (\text{total revenue per item} - \text{manufacturing cost per item} \\ - \text{expected future cost per item}). \tag{4.9}$$

Mathematically, we write this as:

$$\max z(w,x) = r(1 - e^{-0.1w}) - c\left(\frac{w\pi x^2}{4}\right) - \mathscr{Q}(w,x) \\ \text{s. t. } 0 \le w \le w^{\max}, 0 \le x \le x^{\max} \ . \tag{4.10}$$

In stochastic programming terms, this formulation gives the deterministic equivalent problem to the stochastic program for minimizing the current value for the design decision plus future reactions to deviations in the axle diameter. Standard optimization procedures can be used to solve this problem. Assuming maximum values of $w^{\max} = 36$, $x^{\max} = 1.25$, a maximum sales price of $\$10$ ($r = 10$), a material cost of $\$0.025$ per cubic inch ($c = .025$), and a unit penalty $q = 1$, an optimal solution is found at $w^* = 33.6$, $x^* = 1.038$, and $z^* = z(w^*, x^*) = 8.94$. The graphs of z as a function of w for $x = x^*$ and as a function of x for $w = w^*$ appear in Figures 8 and 9. In this solution, the stress constraint is only violated when $.9x = 0.934 \le \xi \le 0.949 = (w/39.27)^{1/3}$.

We again consider the expected value problem where random variables are replaced with their means to obtain a deterministic problem. For this problem, we would obtain:

$$\max z(w,x,\bar{\xi}) = r(1 - e^{-0.1w}) - c\left(\frac{w\pi x^2}{4}\right) \\ - q[\max\{0, \left(\frac{w}{x^3}\right) - 39.27, \left(\frac{w^3}{300x^4}\right) - 210.56\}]^2 \\ \text{s. t. } 0 \le w \le w^{\max}, \quad 0 \le x \le x^{\max} \ . \tag{4.11}$$

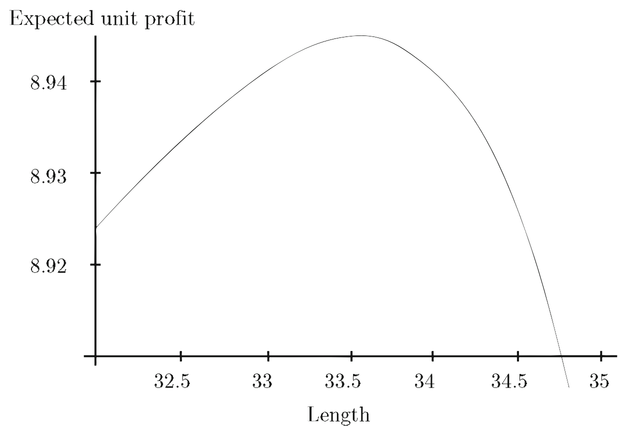

Fig. 8 The expected unit profit as a function of length with a diameter of 1.038 inches.

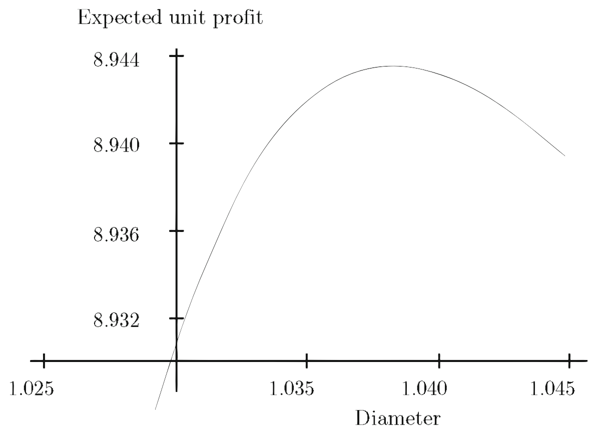

Fig. 9 The expected unit profit as a function of diameter with a length of 33.6 inches.

Using the same data as earlier, an optimal solution to (4.11) is $\bar{w}(\bar{\xi}) = 35.0719$, $\bar{x}(\bar{\xi}) = 0.963$, and $z(\bar{w}, \bar{x}, \bar{\xi}) = 9.07$.

At first glance, it appears that this solution obtains a better expected profit than the stochastic problem solution. However, as we shall see in Chapter 8 on approximations, this deterministic problem paints an overly optimistic picture of the actual situation. The deterministic objective is (in the case of concave maximization) always an *overestimate* of the actual expected profit. In this case, the true *expected* value of the deterministic solution is $z(\bar{w}, \bar{x}) = -26.8$. This problem then has a value of the stochastic solution equal to the difference between the expected value of the stochastic solution and the expected value of the deterministic solution,

$z^* - z(\bar{w}, \bar{x}) = 35.7$. In other words, solving the stochastic program results in a significant profit compared to a considerable loss associated with solving the deterministic problem.

This problem is another example of how stochastic programming can be used. The problem has nonlinear functions and a simple recourse structure. We will discuss further computational methods for problems of this type in Chapter 5. In other problems, decisions may also be taken after the observation of the outcome. For example, we could inspect and then decide whether to sell the product (Exercise 3). This often leads to tolerance settings and is the focus of much of quality control.

The general stochastic program provides a framework for uniting design and quality control. Many loss functions can be used to measure performance degradation to help improve designs in their initial stages. These functions may include the stress and performance penalties described earlier, the Taguchi-type quadratic loss, or methods based on reliability characterizations.

Most traditional approaches assume some form for the distribution as we have done here. This situation rarely matches practice, however. Approximations can nevertheless be used that obtain bounds on the actual solution value so that robust decisions may be made without complete distributional information. This topic will be discussed further in Chapter 8.

Exercises

1. For the example given, what is the probability of exceeding the stress constraint for an axle designed according to the stochastic program optimal specifications?
2. Again, for the example given, what is the probability of exceeding the stress constraint for an axle designed according to the deterministic program's (4.11) optimal specifications?
3. Suppose that every axle can be tested before being shipped at a cost of s per test. The test completely determines the dimensions of the product and thus informs the producer of the risk of failure. Formulate the new problem with testing.

1.5 A Routing Example

a. Presentation

Consider the following simplified vehicle routing problem. A vehicle has to visit four clients (A, B, C, D) in a route starting and ending at a depot (or at the "home sweet home" of the traveling salesperson). One single vehicle of capacity 10 is available. There is no limit on the travel time, so that the vehicle can make consecutive legs if needed.

It is easy to represent a routing problem on a graph (see Figure 10.). A graph $G=(V,E)$ consists of a set V of vertices (or nodes) and a set E of edges (or arcs). Here, the nodes correspond to the set of clients plus the depot $V=\{0,A,B,C,D\}$ where 0 is the depot. Arc (i,j) corresponds to traveling from node i to node j. Arcs may be traveled in either direction. We assume that the vehicle can travel from any point (client or depot) to another. This is equivalent to saying that the graph is complete.

The demands of clients A, B and D are known and equal to 2. Demand of client C is random. To put things to the extreme, assume that the demand of C is either 1 or 7 with equal probability $\frac{1}{2}$. (As we will see later, the example also works with less extreme situations, like a demand of 3 and 5 with equal probability. Direct calculation of all cases is easier here as there are more infeasible cases). All demands must be served. To make things clear, we assume in the sequel that demand is collected at the client. All results and terminologies are easily adapted if demand is delivered. The case of simultaneous pick-ups and deliveries is more involved.

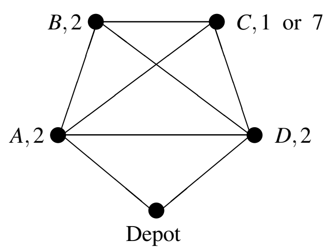

Fig. 10 Graph representation of the vehicle routing problem.

The distances between any two points are given under the form of a symmetrical matrix $C=(c_{ij})$, where c_{ij} is the distance between i and j. Data are in Table 7.

Table 7 Distance matrix.

	0	A	B	C	D
0	–	2	4	4	1
A	2	–	3	4	2
B	4	3	–	1	3
C	4	4	1	–	3
D	1	2	3	3	–

The distance matrix is symmetrical, which means that the distance between two points is the same when traveling in either direction. Distance matrices usually satisfy the so-called *triangle inequality*:

$$c_{ij} \leq c_{ik} + c_{kj} \qquad \forall i, j, k \; . \tag{5.1}$$

The triangle inequality simply means that it is shorter (or at least not longer) to go directly from i to j than through an intermediate node k. The distance matrix in Table 7 satisfies the triangle inequality, but not always strictly. As an example, the distance between A and C is equal to the distance between A and B plus that between B and C. This is due to using small integer data.

The problem of finding the shortest route to visit all clients starting and ending at the depot is known as the TSP (traveling salesperson problem). The optimal TSP route is $(0,A,B,C,D,0)$ of length 10.

This is checked by using a TSP solver. This can also be checked by brute force calculation of all routes. For a problem with n clients, there are $n!$ routes. Indeed, starting from the depot, there are n possible clients to be visited first. When the first client is fixed, there remain $(n-1)$ clients to be visited next and so on. By symmetry, only half of the $n!$ routes have to be checked. As an example, $(0,D,C,B,A,0)$ has the same length as $(0,A,B,C,D,0)$. Here, 12 routes have to be checked. Alternatively, you may trust the authors.

Finding the shortest distance or TSP route is not enough here: the vehicle has a limited capacity of 10 and the demand at C is random. The treatment of the uncertainty depends on the moment when the information becomes available.

b. Wait-and-see solutions

A first case is when the level of the demand is known before starting the route. This could be the case, for instance, if the delivered product is part of a just-in-time production process. If the process works in batches, the number of batches required in C may be 1 or 7, depending on the production process. But the number of batches may then be adequately forecasted.

Alternatively, the products may be wastes generated during the production process. The amount to be collected can be known if an agreement exists with the client or if the client is a subsidiary.

This is known as a situation of *a priori information*. The decision process corresponds to the *wait-and-see* approach. It consists of making the choice of the route after getting the information on the demand level.

The optimal solution in the wait-and-see situation is illustrated in Figure 11.

- Whenever client C requires a single unit to be collected, the vehicle's capacity is large enough to accommodate the demand of the four clients. It is optimal to follow the TSP route of length 10.
- Whenever client C requires 7 units, the total demand of 13 exceeds the vehicle's capacity. The vehicle must travel two successive routes. The combination of

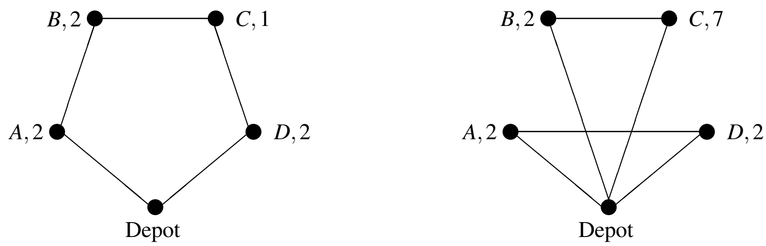

Fig. 11 Wait-and-see solutions (when demand in C is 1 or 7).

two routes with smallest distance is the sequence $(0,A,D,0,B,C,0)$ of total distance 14.

This can be checked as follows. As the demand of C is 7 and the vehicle capacity is 10, the part of the route that visits C can either visit C alone or C with one other client.

There are three possibilities in the first case depending on the order of visit of A, B and D, the best one being $(0,A,B,D)$. There are also three possibilities for the second case, depending on the client which belongs to the route visiting C.

As both situations occur half of the time, optimal routes of length 10 and 14 are traveled half time each. It follows that the mean (or expected) distance traveled under the wait-and-see approach is

$$WS = \frac{1}{2}\ 10 + \frac{1}{2}\ 14 = 12\ .$$

c. Expected value solution

If the demand is not known in advance, it is discovered when arriving at client C. One first attitude is to forget uncertainty. The route is planned in view of the expected demand. As the expected demand of client C is 4, the vehicle's capacity is large enough to accommodate the demand of the four clients (in fact, the expected demand of C and the known demand of the other clients). It is optimal to follow the TSP route $(0,A,B,C,D,0)$ of length 10.

Planning for the expected case is in fact "forgetting" uncertainty. It does not mean uncertainty is absent. To say it in other words, "even if you forget uncertainty, uncertainty will not forget you".

Demand in C is revealed when arriving in C. It is 1 half of the time and 7 the other half of the time, but in a random fashion. Figure 12 shows what really happens.

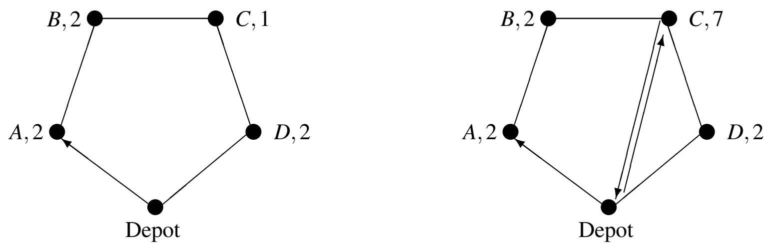

Fig. 12 Effective travel (when demand in C is 1 or 7) if TSP route is planned.

- When the vehicle arrives in C and demand is 1 , it simply proceeds with the planned route. The total demand is 7 and is less than the capacity. The traveled distance is 10. Everything goes well in a beautiful world.
- When the vehicle arrives in C , its load is already 4 . If the demand in C is 7 , the vehicle is unable to collect the total demand. Assuming the goods are divisible, it collects 6 units, then returns to the depot to unload, goes back to C to take the last unit and resumes its trip. The vehicle travels $(0,A,B,C,0,C,D,0)$ for a total length of 18 . In the routing literature, the situation when a vehicle is unable to load a client's demand is known as a *failure*. The extra distance traveled due to this failure is a *return trip* to the depot. The length of 18 is equal to the planned distance 10 of the TSP tour plus the distance 8 of the return trip from C . You may also observe that the same solution is obtained if goods are not divisible.

As both situations occur half of the time, the true cost under uncertainty of the expected value solution is the so-called *expectation of the expected value problem* or

$$EEV = \frac{1}{2}\,10 + \frac{1}{2}\,18 = 14\ .$$

d. Recourse solution

Let us now improve the route choice, in view of the uncertainty at C .

First, observe that it is possible to travel the TSP route $(0,A,B,C,D,0)$ in the opposite direction. The situation is represented on Figure 13. Travelling $(0,D,C,B,A,0)$ implies that

- when the vehicle arrives in C and demand is 1 , it simply proceeds with the planned route. The traveled distance is 10 , as before.
- when the vehicle arrives in C and demand is 7 , the vehicle is able to collect the demand in C . It will not be able to collect the total demand. After collecting demand in C , it returns to the depot, unloads, and then goes to B and A . This

situation is known as a *preventive return*. (It is already known in C that the load in B cannot be collected. It is thus better to return to the depot and resume the tour in B, instead of going to B and making a return trip to the depot.) The vehicle travels $(0,D,C,0,B,A,0)$ for a total length of 17.

The true cost under uncertainty of traveling $(0,D,C,B,A,0)$ is $\frac{1}{2}\,10+\frac{1}{2}\,17=13.5$.

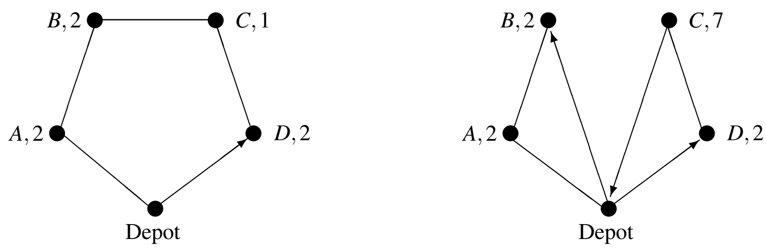

Fig. 13 Effective travel (when demand in C is 1 or 7) if TSP route is planned counterclockwise.

Thus, we have seen that the uncertainty implies that there is a difference between a planned route and the route that is effectively traveled. In the stochastic terminology, deciding on the planned route (or a priori route) is a *first-stage decision*, taken before the random parameters are known. When the uncertainty is revealed, additional or second stage actions are possible. They are called *recourse actions*. In the present example, we have two possible such actions: a return trip to the depot or a preventive return.

After some calculations, it turns out that the optimal solution is to select $(0,C,B,A,D,0)$ as the planned route. If demand in C is 1, the route is followed with length 11. Otherwise, a preventive return occurs in B. The traveled route is $(0,C,B,0,A,D,0)$ with length 14. The optimal solution is represented in Figure 14. The expected length under the optimal recourse policy is

$$RP=\frac{1}{2}\,11+\frac{1}{2}\,14=12.5\,.$$

This example illustrates three important aspects of stochastic programming:

- when dealing with uncertainty, it is important to consider what happens before (first-stage) and after (second-stage) the uncertainty is revealed. It is also important to consider a wider variety of decisions (reversing the travel direction in the first-stage, or doing return trips or preventive returns in the second-stage in this example).
- due to uncertainty, a worse solution is often chosen in the favorable case. This happens here. When demand is low, the vehicle travels the planned route $(0,C,B,A,D,0)$, which is longer than the TSP tour. This may seem stupid: “why

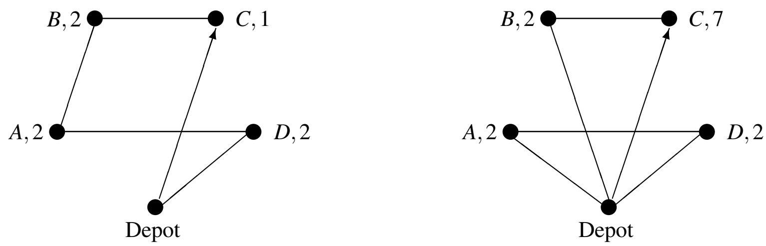

Fig. 14 Effective travel (when demand in C is 1 or 7) if optimal recourse route is planned.

didn't you simply pick up the shortest route?" or lead to some "regret." The reason is simple. By visiting C first, the demand becomes known early in the route and an efficient recourse action (preventive return after B) can be taken when the demand in C is high. This implies indeed some extra cost when the demand in C is low.

- the following relations hold :

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

The first relation $WS \leq RP$ simply says that it is always better to get the information in advance. The difference $RP - WS$ is known as the $EVPI$, expected value of perfect information. Here, $EVPI = 0.5$. This is the maximal amount the planner would be ready to pay client C to get the information in advance. The second relation says that it is better to solve the stochastic program than to pretend uncertainty does not exist. The difference $EEV - RP$ is known as the VSS , value of stochastic solution. Here, $VSS = 1.5$. It tells says that dealing with uncertainty really matters.

e. Other random variables

The present example may seem a bit extreme, with a demand being either 1 or 7 . In fact, it extends to more general random variables. Let $\boldsymbol{\xi}$ denote the random demand in C . We assume $\boldsymbol{\xi}$ has an expectation of 4 (as above). We also assume that the probability of a negative demand is negligible and, similarly, that the probability of $\boldsymbol{\xi}$ exceeding 8 is negligible.

Denote by $p_f = \mathrm{P}(\boldsymbol{\xi} > 4)$, where the index f is a mnemonic to recall that a failure will occur if the expected value solution is chosen. Then the following relations hold:

$$WS = (1 - p_f)10 + p_f 14 \ ,$$

$$EEV = (1-p_f)10 + p_f 18 ,$$
$$RP = (1-p_f)11 + p_f 14 .$$

In the wait-and-see case, the TSP route of length 10 is optimal when demand is less than or equal to 4 and the sequence $(0,A,D,0,B,C,0)$ with length 14 otherwise. In the EEV , a distinction is made between no failure (length 10) or a failure with a return trip (length 18). Finally, in the RP , the route is either $(0,C,B,A,D,0)$ with length 11 when demand is less or equal to 4 or $(0,C,B,0,A,D,0)$ with length 14 otherwise.

Now, consider that demand in C follows a normal distribution with expectation 4 and a variance such that $\mathrm{P}(\boldsymbol{\xi} < 0) \cong 0$. Symmetry implies $\mathrm{P}(\boldsymbol{\xi} > 8) \cong 0$. Symmetry also implies $p_f = \frac{1}{2}$. Thus, all results obtained in the above discrete case are also obtained in the same manner for a normal distribution. The same is true for any continuous uniform distribution of the type $\boldsymbol{\xi} \sim U[4-a, 4+a]$, with $0 < a \leq 4$.

The table of the Poisson(4) distribution shows that $p_f = 0.371$. However, there exists a nonzero probability of the demand exceeding 8 . We may denote this probability as $p_e = \mathrm{P}(\boldsymbol{\xi} > 8) = 0.0214$. If demand exceeds 8 , the recourse solution must be adapted as traveling $(0,B,C,0)$ becomes infeasible. A possible solution for the recourse case is to travel $(0,C,B,D,A,0)$ with length 11 when demand is less or equal to 4 , travel $(0,C,B,0,A,D,0)$ with length 14 when demand is between 5 and 8 and, finally, travel $(0,C,0,A,B,D,0)$ with length 17 otherwise. The corresponding expected cost is:

$$\text{Expected cost } = (1-p_f)11 + (p_f - pe)14 + p_e 17 .$$

f. Chance-constraints

The chance-constraint approach consists of finding the smallest distance feasible route or sequence of routes. A route or sequence of routes is feasible if the vehicle can collect the total demand with a large probability. A typical large probability is, as usual, 90 or 95%. To make things concrete, we take a 95% requirement. This corresponds to a 5% probability of failure.

In the initial example, demand is 1 or 7 with probability $\frac{1}{2}$. Feasibility with a 95% confidence level implies demand of 7 must always be collected. If not, the confidence of the solution would only be 50%. The chance-constraint solution is the sequence $(0,A,D,0,B,C,0)$ of total distance 14 , much worse than the recourse solution.

In line with the previous subsection, we now show how to deal with other random variables.

Let $\boldsymbol{\xi}$ be the random variable representing the demand in C . Any route that does not return to the depot has a capacity of 10 . The probability that it can cover the demand is equal to $\mathrm{P}(6+\boldsymbol{\xi} \leq 10) = \mathrm{P}(\boldsymbol{\xi} \leq 4) = 1 - p_f$. Any route that returns once to the depot consists of two legs, each having a capacity of 10 . Feasibility depends on the leg that visits C (as the other leg has a known demand less than the vehicle capacity).

We can summarize all cases as follows:

- visiting C with the three other clients is feasible with probability $\mathrm{P}(\boldsymbol{\xi} \leq 4) = 1 - p_f$. The best such route is the TSP tour $(0,A,B,C,D,0)$ of length 10.
- visiting C with two other clients is feasible with probability $\mathrm{P}(4 + \boldsymbol{\xi} \leq 10) = \mathrm{P}(\boldsymbol{\xi} \leq 6)$. The smallest distance corresponding route is the sequence $(0,D,0,A,B,C,0)$ of total distance 12.
- if C is visited with one other client, the route is feasible with probability $\mathrm{P}(\boldsymbol{\xi} \leq 8)$. The corresponding route with smallest distance is the sequence $(0,A,D,0,B,C,0)$ of total distance 14.
- if the leg that visits C does not visit any other client, it is feasible with probability $\mathrm{P}(\boldsymbol{\xi} \leq 10)$. The best corresponding route is $(0,C,0,A,B,D,0)$ of length 17.

The various solutions have increased lengths but also increased probabilities of being feasible. To find the chance-constraint solution, it suffices to consider each case in turn. The first that has a probability larger than the requested 95% is the chance-constraint solution.

For a Poisson random variable with expectation 4, $p_f = 0.371$ and thus any route that does not return to the depot is infeasible. A route that returns once to the depot and visits C with two other clients has a probability $\mathrm{P}(\boldsymbol{\xi} \leq 6) = 0.8893$ to cover the demand and is thus infeasible. A route that returns once to the depot and visits C with at most one other client has a probability $\mathrm{P}(\boldsymbol{\xi} \leq 8) = 0.9786$ to cover the demand. The route $(0,A,D,0,B,C,0)$ is, as before, the optimal solution for a 95% chance-constraint.

Exercises

1. Consider a continuous uniform distribution of the type $\boldsymbol{\xi} \sim U[4-a, 4+a]$, with $0 < a \leq 4$. Obtain the optimal chance constraint solution as a function of a.
2. Consider the case where the demand in C follows a Normal distribution with expectation 4 and a variance such that $\mathrm{P}(\boldsymbol{\xi} < 0) \cong 0$. Obtain the optimal chance constraint solution as a function of σ.

1.6 Other Applications

In this chapter, we discussed a few examples of stochastic programming applications. The examples were chosen because of their frequency in stochastic programming application as well as to illustrate various aspects of stochastic programming models in terms of number of stages, continuous or discrete variables, separable or nonseparable recourse, probabilistic constraints, and linear or nonlinear constraint and objective functions.

Several other application areas deserve some recognition but were not discussed yet. A particular example is in airline planning. One of the first applications of stochastic programming was a decision on the allocation of aircraft to routes (*fleet assignment*) by Ferguson and Dantzig [1956]. In this problem, penalties were incurred for lost passengers. The problem becomes a simple recourse problem in stochastic programming terms that they solved using a variant of the standard transportation simplex method (see Section 5.7).

Production planning is another major area that was not in our examples. This area also has been the subject of stochastic programming models for many years. The original chance-constrained stochastic programming model of Charnes, Cooper, and Symonds [1958], for example, considered the production of heating oil with constraints on meeting sales and not exceeding capacity. Other examples include the study by Escudero et al. [1993] for IBM procurement policies.

Water resource modeling has also received widespread application. A good example of this area is the paper by Prékopa and Szántai [1976], where they discuss regulation of Lake Balaton's water level and show how stochastic programming could have avoided floods that occurred before such planning methods were available. Approaches to pollution and the environmental area of water resource planning are also common. An example discussion appears in Somlyódy and Wets [1988].

Energy planning has been the focus of many stochastic programming studies. We note in particular Manne's [1974] analysis of the U.S. decision on whether to invest in breeder reactors. The more recent work of Manne and Richels [1992] on buying insurance against the greenhouse effect is also an excellent example of how stochastic programming can model uncertain future situations so that informed public policy decisions may be made.

Stochastic programming has been applied in many other areas. Of particular note is the forestry planning model in Gassmann ([1989]) and the hospital staffing problem in Kao and Queyranne ([1985]). We also include two exercises in stochastic programming in sports. Many other references appear in King's survey (King [1988b]), the volume by Ermoliev and Wets [1988], and the collection edited by Wallace and Ziemba [2005]. Many more applications are open to stochastic programming, especially with the powerful techniques now available. In the remainder of this book, we will explore those methods, their properties, and the general classes of problems they solve.

Exercises

These exercises all contain a stochastic programming problem that can be solved using standard linear, nonlinear and integer programming software. For each problem, you should develop the model, solve the stochastic program, solve the expected value problem, and find the value of the stochastic solution.

1. Northam Airlines is trying to decide how to partition a new plane for its Chicago–Detroit route. The plane can seat 200 economy class passengers. A section can be partitioned off for first class seats but each of these seats takes the space of 2 economy class seats. A business class section can also be included, but each of these seats takes as much space as 1.5 economy class seats. The profit on a first class ticket is, however, three times the profit of an economy ticket. A business class ticket has a profit of two times an economy ticket's profit. Once the plane is partitioned into these seating classes, it cannot be changed. Northam knows, however, that the plane will not always be full in each section. They have decided that three scenarios will occur with about the same frequency: (1) weekday morning and evening traffic, (2) weekend traffic, and (3) weekday midday traffic. Under Scenario 1, they think they can sell as many as 20 first class tickets, 50 business class tickets, and 200 economy tickets. Under Scenario 2, these figures are 10, 25, and 175. Under Scenario 3, they are 5, 10, and 150. You can assume they cannot sell more tickets than seats in each of the sections. (In reality, the company may allow overbooking, but then it faces the problem of passengers with reservations who do not appear for the flight (*no-shows*). The problem of determining how many passengers to accept is part of the field called *yield management* or *revenue management*. For one approach to this problem, see Brumelle and McGill [1993]. This subject is explored further in Exercise 1 of Section 2.7.)
2. Tomatoes Inc. (TI) produces tomato paste, ketchup, and salsa from four resources: labor, tomatoes, sugar, and spices. Each box of the tomato paste requires 0.5 labor hours, 1.0 crate of tomatoes, no sugar, and 0.25 can of spice. A ketchup box requires 0.8 labor hours, 0.5 crate of tomatoes, 0.5 sacks of sugar, and 1.0 can of spice. A salsa box requires 1.0 labor hour, 0.5 crate of tomatoes, 1.0 sack of sugar, and 3.0 cans of spice.

   The company is deciding production for the next three periods. It is restricted to using 200 hours of labor, 250 crates of tomatoes, 300 sacks of sugar, and 100 cans of spices in each period at regular rates. The company can, however, pay for additional resources at a cost of 2.0 per labor hour, 0.5 per tomato crate, 1.0 per sugar sack, and 1.0 per spice can. The regular production costs for each product are 1.0 for tomato paste, 1.5 for ketchup, and 2.5 for salsa.

   Demand is not known with certainty until after the products are made in each period. TI forecasts that in each period two possibilities are equally likely, corresponding to a good or bad economy. In the good case, 200 boxes of tomato paste, 40 boxes of ketchup, and 20 boxes of salsa can be sold. In the bad case, these values are reduced to 100, 30, and 5, respectively. Any surplus production is stored at costs of 0.5, 0.25, and 0.2 per box for tomato paste, ketchup, and salsa, respectively. TI also considers unmet demand important and assigns costs of 2.0, 3.0, and 6.0 per box for tomato paste, ketchup, and salsa, respectively, for any demand that is not met in each period.
3. The Clear Lake Dam controls the water level in Clear Lake, a well-known resort in Dreamland. The Dam Commission is trying to decide how much water to release in each of the next four months. The Lake is currently 150 mm below flood

stage. The dam is capable of lowering the water level 200 mm each month, but additional precipitation and evaporation affect the dam. The weather near Clear Lake is highly variable. The Dam Commission has divided the months into two two-month blocks of similar weather. The months within each block have the same probabilities for weather, which are assumed independent of one another. In each month of the first block, they assign a probability of 1/2 to having a natural 100-mm increase in water levels and probabilities of 1/4 to having a 50-mm decrease or a 250-mm increase in water levels. All these figures correspond to natural changes in water level without dam releases. In each month of the second block, they assign a probability of 1/2 to having a natural 150-mm increase in water levels and probabilities of 1/4 to having a 50-mm increase or a 350-mm increase in water levels. If a flood occurs, then damage is assessed at $10,000 per mm above flood level. A water level too low leads to costly importation of water. These costs are $5000 per mm less than 250 mm below flood stage. The commission first considers an overall goal of minimizing expected costs. They also consider minimizing the probability of violating the maximum and minimum water levels. (This makes the problem a special form of chance-constrained model.) Consider both objectives.

4. The Energy Ministry of a medium-size country is trying to decide on expenditures for new resources that can be used to meet energy demand in the next decade. There are currently two major resources to meet energy demand. These resources are, however, exhaustible. Resource 1 has a cost of 5 per unit of demand met and a total current availability equal to 25 cumulative units of demand. Resource 2 has a cost of 10 per unit of demand met and a total current availability of 10 demand units. An additional resource from outside the country is always available at a cost of 16.7 per unit of demand met.

   Some investment is considered in each of Resources 1 and 2 to discover new supplies and build capital. Resource 1 is, however, elusive. A unit of investment in new sources of Resource 1 yields only 0.1 demand unit of Resource 1 with probability 0.5 and yields 1 demand unit with probability 0.5 . For Resource 2, investment is well known. Each unit of investment yields a demand unit equivalent of Resource 2. Cumulative demand in the current decade is projected to be 10 , while demand in the next decade will be 25 .

   The ministry wants to minimize expected costs of meeting demands in the current and following decade assuming that the results of Resource 1 investment will only be known when the current decade ends. Next-decade costs are discounted to 60% of their future real values (which should not change).

5. Pacific Pulp and Paper is deciding how to manage their main forest. They have trees at a variety of ages, which we will break into Classes 1 to 4 . Currently, they have 8000 acres in Class 1 , 10,000 acres in Class 2 , 20,000 in Class 3, and 60,000 in Class 4 . Each class corresponds to about 25 years of growth. The company would like to determine how to harvest in each of the next four 25-year periods to maximize expected revenue from the forest. They also foresee the company's continuing after a century, so they place a constraint of having 40,000 acres in Class 4 at the end of the planning horizon.

Each class of timber has a different yield. Class 1 has no yield, Class 2 yields 250 cubic feet per acre, Class 3 yields 510 cubic feet per acre, and Class 4 yields 700 cubic feet per acre. Without fires, the number of acres in Class i (for $i = 2, 3$) in one period is equal to the amount in Class $i - 1$ from the previous period minus the amount harvested from Class $i - 1$ in the previous period. Class 1 at period t consists of the total amount harvested in the previous period $t - 1$, while Class 4 includes all remaining Class 4 land plus the increment from Class 3.

While weather effects do not vary greatly over 25-year periods, fire damage can be quite variable. Assume that in each 25-year block, the probability is 1/3 that 15% of all timber stands are destroyed and that the probability is 2/3 that 5% is lost. Suppose that discount rates are completely overcome by increasing timber value so that all harvests in the 100-year period have the same current value. Revenue is then proportional to the total wood yield.

6. A hospital emergency room is trying to plan holiday weekend staffing for a Saturday, Sunday, and Monday. Regular-time nurses can work any two days of the weekend at a rate of \$300 per day. In general, a nurse can handle 10 patients during a shift. The demand is not known, however. If more patients arrive than the capacity of the regular-time nurses, they must work overtime at an average cost of \$50 per patient overload. The Saturday demand also gives a good indicator of Sunday–Monday demand. More nurses can be called in for Sunday–Monday duty after Saturday demand is observed. The cost is \$400 per day, however, in this case. The hospital would like to minimize the expected cost of meeting demand.

   Suppose that the following scenarios of 3-day demand are all equally likely: $(100, 90, 20)$, $(100, 110, 120)$, $(100, 100, 110)$, $(90, 100, 110)$, $(90, 80, 110)$, $(90, 90, 100)$, $(80, 90, 100)$, $(80, 70, 100)$, and $(80, 80, 90)$.

7. After winning the pole at Monza, you are trying to determine the quickest way to get through the first right-hand turn, which begins 200 meters from the start and is 30 meters wide. You are through the turn at 100 meters past the beginning of the next stretch (see Figure 15). As in the figure, you will attempt to stay 10 meters inside the barrier on the starting stretch (maintaining this distance from each barrier as accelerate as fast as possible until point d_1 . At this distance, you will start braking as hard as possible and take the turn at the current velocity reached at some point d_2 . (Assume a circular turn with radius equal to the square of velocity divided by maximum lateral acceleration.) Obviously, you do not want to go off the course.

   The problem is that you can never be exactly sure of the car and track speed until you start braking at point d_1 . At that point, you can tell whether the track is fast, medium, or slow, and you can then determine the point d_2 where you enter the turn. You suppose that the three kinds of track/car combinations are equally likely. If fast, you accelerate at 27 m/sec ^{2} , decelerate at 45 m/sec ^{2} , and have a maximum lateral acceleration of 1.8 g (= 17.5 m/sec ^{2}). For medium, these values are 24 , 42 , and 16 ; for slow, the values are 20 , 35 , and 14 . You want to minimize the expected time through this section. You also assume that

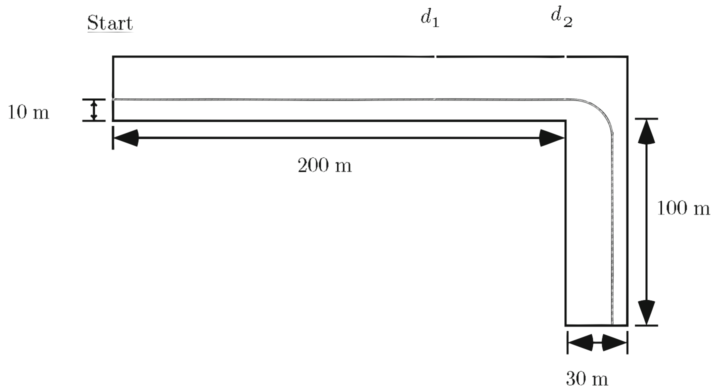

Fig. 15 Opening straight and turn for Problem 7.

if you follow an optimal strategy, other competitors will not throw you out of the race (although you may not be sure of that). After finding the optimal strategy for any feasible position on the second straight-away, find an optimal strategy with a constraint to remain no more than 10 meters from the inside wall after completing the turn and compare the results.

8. In training for the Olympic decathlon, you are trying to choose your takeoff point for the long jump to maximize your expected official jump. Unfortunately, when you aim at a certain spot, you have a 50/50 chance of actually taking off 10 cm beyond that point. If that violates the official takeoff line, you foul and lose that jump opportunity. Assume that you have three chances and that your longest jump counts as your official finish.

   You then want to determine your aiming strategy for each jump. Assume that your actual takeoff is independent from jump to jump. Initially you are equally likely to hit a 7.4- or 7.6-meter jump from your actual takeoff point. If you hit a long first jump, then you have a 2/3 chance of another 7.6-meter jump and 1/3 chance of jumping 7.4 meters. The probabilities are reversed if you jumped 7.4 meters the first time. You always seem to hit the third jump the same as the second.

   First, find a strategy to maximize the expected official jump. Then, maximize decathlon points from the following Table 8.

Table 8 Decathlon Points for Problem 8.

Distance	Points	Distance	Points
7.30	886	7.46	925
7.31	888	7.47	927
7.32	891	7.48	930
7.33	893	7.49	932
7.34	896	7.50	935
7.35	898	7.51	937
7.36	900	7.52	940
7.37	903	7.53	942
7.38	905	7.54	945
7.39	908	7.55	947
7.40	910	7.56	950
7.41	913	7.57	952
7.42	915	7.58	955
7.43	918	7.59	957
7.44	920	7.60	960
7.45	922	7.61	962

Chapter 2
Uncertainty and Modeling Issues

In the previous chapter, we gave several examples of stochastic programming models. These formulations fit into different categories of stochastic programs in terms of the characteristics of the model. This chapter presents those basic characteristics by describing the fundamentals of any modeling effort and some of the standard forms detailed in later chapters.

Before beginning general model descriptions, however, we first describe the probability concepts that we will assume in the rest of the book. Familiarity with these concepts is essential in understanding the structure of a stochastic program. This presentation is made simple enough to be understood by readers unfamiliar with the field and, thus, leaves aside some questions related to measure theory. Sections 2.2 through 2.7 build on these fundamentals and give the general forms in various categories. Section 2.8 provides a detailed discussion of a modeling exercise. Sections 2.9 and 2.10 give alternative characterizations of stochastic optimization problems and some background on the relationship of stochastic programming to other areas of decision making under uncertainty. Section 2.11 briefly reviews the main optimization concepts used in the book.

2.1 Probability Spaces and Random Variables

Several parameters of a problem can be considered uncertain and are thus represented as random variables. Production and distribution costs typically depend on fuel costs, which are random. Future demands depend on uncertain market conditions. Crop returns depend on uncertain weather conditions.

Uncertainty is represented in terms of random experiments with outcomes denoted by ω. The set of all outcomes is represented by Ω. In a transport and distribution problem, the outcomes range from political conditions in the Middle East to general trade situations, while the random variable of interest may be the fuel cost. The relevant set of outcomes is clearly problem-dependent. Also, it is usually not

J.R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-1-4614-0237-4_2,

© Springer Science+Business Media, LLC 2011

very important to be able to define those outcomes accurately because the focus is mainly on their impact on some (random) variables.

The outcomes may be combined into subsets of Ω called *events*. We denote by $\mathscr{A}$ a collection of random events. As an example, if Ω contains the six possible results of the throw of a die, $\mathscr{A}$ also contains combined outcomes such as an odd number, a result smaller than or equal to four, etc. If Ω contains weather conditions for a single day, $\mathscr{A}$ also contains combined events such as "a day without rain," which might be the union of a sunny day, a partly cloudy day, a cloudy day without showers, etc.

Finally, to each event $A \in \mathscr{A}$ is associated a value $\mathrm{P}(A)$, called a *probability*, such that $0 \le \mathrm{P}(A) \le 1$, $\mathrm{P}(\emptyset) = 0$, $\mathrm{P}(\Omega) = 1$ and $\mathrm{P}(A_1 \cup A_2) = \mathrm{P}(A_1) + \mathrm{P}(A_2)$ if $A_1 \cap A_2 = \emptyset$. The triplet $(\Omega, \mathscr{A}, P)$ is called a *probability space* that must satisfy a number of conditions (see, e.g., Chung [1974]). It is possible to define several random variables associated with a probability space, namely, all variables that are influenced by the random events in $\mathscr{A}$. If one takes as elements of Ω events ranging from the political situation in the Middle East to the general trade situations, they allow us to describe random variables such as the fuel costs and the interest rates and inflation rates in some Western countries. If the elements of Ω are the weather conditions from April to September, they influence random variables such as the production of corn, the sales of umbrellas and ice cream, or even the exam results of undergraduate students.

In terms of stochastic programming, there exists one situation where the description of random variables is closely related to Ω: in some cases indeed, the elements $\omega \in \Omega$ are used to describe a few *states of the world* or *scenarios*. All random elements then jointly depend on these finitely many scenarios. Such a situation frequently occurs in strategic models where the knowledge of the possible outcomes in the future is obtained through experts' judgments and only a few scenarios are considered in detail. In many situations, however, it is extremely difficult and pointless to construct Ω and $\mathscr{A}$; the knowledge of the random variables is sufficient.

For a particular random variable $\boldsymbol{\xi}$, we define its cumulative distribution $F_{\boldsymbol{\xi}}(x) = \mathrm{P}(\boldsymbol{\xi} \le x)$, or more precisely $F_{\boldsymbol{\xi}}(x) = \mathrm{P}(\{\omega \mid \boldsymbol{\xi} \le x\})$. Two major cases are then considered. A discrete random variable takes a finite or countable number of different values. It is best described by its probability distribution, which is the list of possible values, ξ^k, $k \in K$, with associated probabilities,

$$f(\xi^k) = \mathrm{P}(\boldsymbol{\xi} = \xi^k) \quad \text{s. t.} \sum_{k \in K} f(\xi^k) = 1 .$$

Continuous random variables can often be described through a so-called *density* function $f(\xi)$. The probability of ξ being in an interval $[a,b]$ is obtained as

$$\mathrm{P}(a \le \boldsymbol{\xi} \le b) = \int_a^b f(\boldsymbol{\xi}) d\boldsymbol{\xi} ,$$

or equivalently

$$P(a \leq \boldsymbol{\xi} \leq b) = \int_a^b dF(\boldsymbol{\xi}) \ ,$$

where $F(\cdot)$ is the cumulative distribution as earlier. Contrary to the discrete case, the probability of a single value $P(\boldsymbol{\xi} = a)$ is always zero for a continuous random variable. The distribution $F(\cdot)$ must be such that $\int_{-\infty}^{\infty} dF(\boldsymbol{\xi}) = 1$.

The *expectation* of a random variable is computed as $\mu = \sum_{k \in K} \xi^k f(\xi^k)$ or $\mu = \int_{-\infty}^{\infty} \boldsymbol{\xi} dF(\boldsymbol{\xi})$ in the discrete and continuous cases, respectively. The *variance* of a random variable is $E[(\boldsymbol{\xi} - \mu)^2]$. The expectation of $\boldsymbol{\xi}^r$ is called the *r th moment* of $\boldsymbol{\xi}$ and is denoted $\bar{\xi}^{(r)} = E[\boldsymbol{\xi}^r]$. A point η is called the α -quantile of $\boldsymbol{\xi}$ if and only if for $0 < \alpha < 1$, $\eta = \min\{x \mid F(x) \geq \alpha\}$.

The appendix lists the distributions used in the textbook and their expectations and variances. The concepts of probability distribution, density, and expectation easily extend to the case of multiple random variables. Some of the sections in the book use probability measure theory which generalizes these concepts. These sections contain a warning to readers unfamiliar with this field.

2.2 Deterministic Linear Programs

A deterministic linear program consists of finding a solution to

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

where x is an $(n \times 1)$ vector of decisions and c , A and b are known data of sizes $(n \times 1)$, $(m \times n)$, and $(m \times 1)$, respectively. The value $z = c^T x$ corresponds to the objective function, while $\{x \mid Ax = b \ , \ x \geq 0\}$ defines the set of feasible solutions. An optimum x^* is a feasible solution such that $c^T x \geq c^T x^*$ for any feasible x . Linear programs typically search for a minimal-cost solution under some requirements (demand) to be met or for a maximum profit solution under limited resources. There exists a wide variety of applications, routinely solved in the industry. As introductory references, we cite Chvátal [1980], Dantzig [1963], and Murty [1983]. We assume the reader is familiar with linear programming and has some knowledge of basic duality theory as in these textbooks. A short review is given in Section 2.11.

2.3 Decisions and Stages

Stochastic linear programs are linear programs in which some problem data may be considered uncertain. *Recourse programs* are those in which some decisions or recourse actions can be taken after uncertainty is disclosed. To be more precise, data uncertainty means that some of the problem data can be represented as random variables. An accurate probabilistic description of the random variables is assumed available, under the form of the probability distributions, densities or, more generally, probability measures. As usual, the particular values the various random variables will take are only known after the random experiment, i.e., the vector $\xi = \xi(\omega)$ is only known after the experiment.

The set of decisions is then divided into two groups:

- A number of decisions have to be taken before the experiment. All these decisions are called *first-stage decisions* and the period when these decisions are taken is called the *first stage*.
- A number of decisions can be taken after the experiment. They are called *second-stage decisions*. The corresponding period is called the *second stage*.

First-stage decisions are represented by the vector x, while second-stage decisions are represented by the vector y or $y(\omega)$ or even $y(\omega,x)$ if one wishes to stress that second-stage decisions differ as functions of the outcome of the random experiment and of the first-stage decision. The sequence of events and decisions is thus summarized as

$$x \to \xi(\omega) \to y(\omega,x) \ .$$

Observe here that the definitions of first and second stages are only related to before and after the random experiment and may in fact contain sequences of decisions and events. In the farming example of Section 1.1, the first stage corresponds to planting and occurs during the whole spring. Second-stage decisions consist of sales and purchases. Selling extra corn would probably occur very soon after the harvest while buying missing corn will take place as late as possible.

A more extreme example is the following. A traveling salesperson receives one item every day. She visits clients hoping to sell that item. She returns home when a buyer is found or when all clients are visited. Clients buy or do not buy in a random fashion. The decision is not influenced by the previous days' decisions. The salesperson wishes to determine the order in which to visit clients, in such a way as to be at home as early as possible (seems reasonable, does it not?). Time spent involves the traveling time plus some service time at each visited client.

To make things simple, once the sequence of clients to be visited is fixed, it is not changed. Clearly the first stage consists of fixing the sequence and traveling to the first client. The second stage is of variable duration depending on the successive clients buying the item or not. Now, consider the following example. There are two clients with probability of buying 0.3 and 0.8, respectively and traveling times (including service) as in the graph of Figure 1.

Assume the day starts at 8 A.M. If the sequence is $(1,2)$, the first stage goes from 8 to 9:30. The second stage starts at 9:30 and finishes either at 11 A.M. if 1 buys or 4:30 P.M. otherwise. If the sequence is $(2,1)$, the first stage goes from 8 to 12:00, the second stage starts at 12:00 and finishes either at 4:00 P.M. or at 4:30 P.M. Thus, the first stage if sequence $(2,1)$ is chosen may sometimes end after the second stage is finished when $(1,2)$ is chosen if Client 1 buys the item.

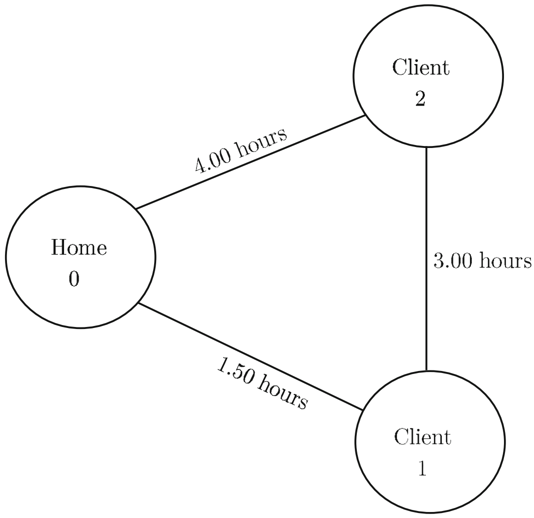

Fig. 1 Traveling salesperson example.

2.4 Two-Stage Program with Fixed Recourse

The classical two-stage stochastic linear program with fixed recourse (originated by Dantzig [1955] and Beale [1955]) is the problem of finding

$$\min z = c^T x + \mathrm{E}_{\boldsymbol{\xi}}[\min q(\omega)^T y(\omega)] \tag{4.1}$$

$$\text{s. t.} \quad Ax = b \,, \tag{4.2}$$

$$T(\omega)x + Wy(\omega) = h(\omega) \,, \tag{4.3}$$

$$x \geq 0 \,, y(\omega) \geq 0 \,. \tag{4.4}$$

As in the previous section, a distinction is made between the first stage and the second stage. The first-stage decisions are represented by the $n_1 \times 1$ vector x. Corresponding to x are the first-stage vectors and matrices c, b, and A, of sizes $n_1 \times 1$, $m_1 \times 1$, and $m_1 \times n_1$, respectively. In the second stage, a number of random events $\omega \in \Omega$ may realize. For a given realization ω, the second-stage problem data $q(\omega)$, $h(\omega)$ and $T(\omega)$ become known, where $q(\omega)$ is $n_2 \times 1$, $h(\omega)$ is $m_2 \times 1$, and $T(\omega)$ is $m_2 \times n_1$.

Each component of q, T, and h is thus a possible random variable. Let $T_{i\cdot}(\omega)$ be the i th row of $T(\omega)$. Piecing together the stochastic components of the second-stage data, we obtain a vector $\xi^T(\omega) = (q(\omega)^T, h(\omega)^T, T_{1\cdot}(\omega), \ldots, T_{m_2\cdot}(\omega))$, with potentially up to $N = n_2 + m_2 + (m_2 \times n_1)$ components. As indicated before, a single random event ω (or state of the world) influences several random variables, here, all components of ξ.

Let also $\Xi \subset \Re^N$ be the *support* of $\boldsymbol{\xi}$, that is, the smallest closed subset in $\Re^N$ such that $\mathrm{P}(\Xi) = 1$. As just said, when the random event ω is realized, the second-stage problem data, q, h, and T, become known. Then, the second-stage decision $y(\omega)$ or $(y(\omega,x))$ must be taken. The dependence of y on ω is of a completely different nature from the dependence of q or other parameters on ω. It is not functional but simply indicates that the decisions y are typically not the same under different realizations of ω. They are chosen so that the constraints (4.3) and (4.4) hold *almost surely* (denoted *a.s.*), i.e., for all $\omega \in \Omega$ except perhaps for sets with zero probability. We assume random constraints to hold in this way throughout this book unless a specific probability is given for satisfying constraints.

The objective function of (4.1) contains a deterministic term $c^T x$ and the expectation of the second-stage objective $q(\omega)^T y(\omega)$ taken over all realizations of the random event ω. This second-stage term is the more difficult one because, for each ω, the value $y(\omega)$ is the solution of a linear program. To stress this fact, one sometimes uses the notion of a deterministic equivalent program. For a given realization ω, let

$$Q(x,\xi(\omega)) = \min_y \{q(\omega)^T y \mid Wy = h(\omega) - T(\omega)x, y \geq 0\} \tag{4.5}$$

be the second-stage value function. Then, define the expected second-stage value function

$$\mathcal{Q}(x) = \mathrm{E}_{\boldsymbol{\xi}} Q(x,\xi(\omega)) \tag{4.6}$$

and the *deterministic equivalent program* (DEP)

$$\min z = c^T x + \mathcal{Q}(x) \tag{4.7}$$

$$\text{s. t. } Ax = b\,, \tag{4.8}$$

$$x \geq 0\,.$$

This representation of a stochastic program clearly illustrates that the major difference from a deterministic formulation is in the second-stage value function. If that function is given, then a stochastic program is just an ordinary nonlinear program.

Formulation (4.1)–(4.4) is the simplest form of a stochastic two-stage program. Extensions are easily modeled. For example, if first-stage or second-stage decisions are to be integers, constraint (4.4) can be replaced by a more general form:

$$x \in X\,, \quad y(w) \in Y\,,$$

where $X = Z_+^{n_1}$ and $Y = Z_+^{n_2}$. Similarly, nonlinear first-stage and second-stage objectives or constraints can easily be incorporated.

Examples of recourse formulation and interpretations

The definition of first stage versus second stage is not only problem dependent but also context dependent. We illustrate different examples of recourse formulations for one class of problems: *the location problem*.

Let $i = 1, \dots, m$ index clients having demand d_i for a given commodity. The firm can open a facility (such as a plant or a warehouse) in potential sites $j = 1, \dots, n$. Each client can be supplied from an open facility where the commodity is made available (i.e., produced or stored). The problem of the firm is to choose the number of facilities to open, their locations, and market areas to maximize profit or minimize costs.

Let us first present the deterministic version of the so-called simple plant location or uncapacitated facility location problem. Let x_j be a binary variable equal to one if facility j is open and zero otherwise. Let c_j be the fixed cost for opening and operating facility j and let v_j be the variable operating cost of facility j. Let y_{ij} be the fraction of the demand of client i served from facility j and t_{ij} be the unit transportation cost from j to i.

All costs and profits should be taken in conformable units, typically on a yearly equivalent basis. Let r_i denote the unit price charged to client i and $q_{ij} = (r_i - v_j - t_{ij})d_i$ be the total revenue obtained when all of client i's demand is satisfied from facility j. Then the simple plant location problem or uncapacitated facility location problem (UFLP) reads as follows:

$$\text{UFLP: } \max_{x,y} z(x,y) = -\sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m}\sum_{j=1}^{n} q_{ij} y_{ij} \tag{4.9}$$

$$\text{s. t. } \sum_{j=1}^{n} y_{ij} \le 1 \,, \qquad i = 1, \dots, m \,, \tag{4.10}$$

$$0 \le y_{ij} \le x_j \,, \qquad i = 1, \dots, m \,,\ j = 1, \dots, n \,, \tag{4.11}$$

$$x_j \in \{0,1\} \,, \qquad j = 1, \dots, n \,. \tag{4.12}$$

Constraints (4.10) ensure that the sum of fractions of clients i's demand served cannot exceed one. Constraints (4.11) ensure that clients are served only through open plants.

It is customary to present the uncapacitated facility location in a different canonical form that minimizes the sum of the fixed costs of opening facilities and of the transportation costs plus possibly the variable operating costs. (There are several ways to arrive at this canonical representation. One is to assume that unit prices are much larger than unit costs in such a way that demand is always fully satisfied.) This presentation more clearly stresses the link between the deterministic and stochastic cases.

In the UFLP, a trade-off is sought between opening more plants, which results in higher fixed costs and lower transportation costs and opening fewer plants with the opposite effect. Whenever the optimal solution is known, the size of an open facility is computed as the sum of demands it serves. (In the deterministic case, it is always optimal to have each y_{ij} equal to either zero or one.) The market areas of each facility are then well-defined.

The notation x_j for the location variables and y_{ij} for the distribution variables is common in location theory and is thus not meant here as first stage and second stage, respectively, although in some of the models it is indeed the case.

Several parameters of the problem may be uncertain and may thus have to be represented by random variables. Production and distribution costs may vary over time. Future demands for the product may be uncertain.

As indicated in the introduction of the section, we will now discuss various situations of recourse. It is customary to consider that the location decisions x_j are first-stage decisions because it takes some time to implement decisions such as moving or building a plant or warehouse. The main modeling issue is on the distribution decisions. The firm may have full control on the distribution, for example, when the clients are shops owned by the firm. It may then choose the distribution pattern after conducting some random experiments. In other cases, the firm may have contracts that fix which plants serve which clients, or the firm may wish fixed distribution patterns in view of improved efficiency because drivers would have better knowledge of the regions traveled.

a. Fixed distribution pattern, fixed demand, r_i, v_j, t_{ij} *stochastic*

Assume the only uncertainties are in production and distribution costs and prices charged to the client. Assume also that the distribution pattern is fixed in advance, i.e., is considered first stage. The second stage then just serves as a measure of the cost of distribution. We now show that the problem is in fact a deterministic problem in which the total revenue $q_{ij} = (r_i - v_j - t_{ij})d_i$ can be replaced by its expectation. To do this, we formally introduce extra second-stage variables w_{ij}, with the constraint $w_{ij}(\omega) = y_{ij}$ for all ω. We obtain

$$\max \; -\sum_{j=1}^{n} c_j x_j + \mathrm{E}_{\boldsymbol{\xi}} \sum_{i=1}^{m} \sum_{j=1}^{n} q_{ij}(\omega) w_{ij}(\omega)$$

s.t. (4.10), (4.11), (4.12), and

$$w_{ij}(\omega) = y_{ij}, \quad i = 1, \ldots, m, \qquad j = 1, \ldots, n \;\; \forall \omega . \tag{4.13}$$

By (4.13), the second-stage objective function can be replaced by

$$\mathrm{E}_{\boldsymbol{\xi}} \sum_{i=1}^{m} \sum_{j=1}^{n} q_{ij}(\omega) y_{ij}$$

or

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{E}_{\boldsymbol{\xi}} q_{ij}(\omega) y_{ij} \ ,$$

because y_{ij} is fixed and summations and expectation can be interchanged. The problem is thus the deterministic problem

$$\max \ -\sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} \sum_{j=1}^{n} (\mathrm{E}_{\boldsymbol{\xi}} q_{ij}(\omega)) y_{ij}$$

s.t. (4.10), (4.11), (4.12).

Although there exists uncertainty about the distribution costs and revenues, the only possible action is to plan in view of the expected costs.

b. Fixed distribution pattern, uncertain demand

Assume now that demand is uncertain, but, for some of the reasons cited earlier, the distribution pattern is fixed in the first stage. Depending on the context, the distribution costs and revenues (v_j, t_{ij}, r_i) may or may not be uncertain.

We define $y_{ij} =$ quantity transported from j to i, a quantity no longer defined as a function of the demand d_i, because demand is now stochastic. For simplicity, we assume that a penalty q_i^+ is paid per unit of demand d_i which cannot be satisfied from all quantities transported to i (they might have to be obtained from other sources) and a penalty q_i^- is paid per unit on the products delivered to i in excess of d_i (the cost of inventory, for example). We thus introduce second-stage variables: $w_i^-(\omega) =$ amount of extra products delivered to i in state ω; $w_i^+(\omega) =$ amount of unsatisfied demand to i in state ω.

The formulation becomes

$$\max - \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} \sum_{j=1}^{n} (\mathrm{E}_{\boldsymbol{\xi}}(-v_j - t_{ij})) y_{ij} + \mathrm{E}_{\boldsymbol{\xi}}[-\sum_{i=1}^{m} q_i^+ w_i^+(\omega)$$

$$- \sum_{i=1}^{m} q_i^- w_i^-(\omega)] + \mathrm{E}_{\boldsymbol{\xi}} \sum_{i=1}^{m} r_i d_i(\omega) \tag{4.14}$$

$$\text{s. t.} \ \sum_{i=1}^{m} y_{ij} \le M x_j \ , \qquad j = 1, \ldots, n \ , \tag{4.15}$$

$$w_i^+(\omega) - w_i^-(\omega) = d_i(\omega) - \sum_{j=1}^{n} y_{ij} \ , \qquad i = 1, \ldots, m \ , \tag{4.16}$$

$$x_j \in \{0,1\} \ , \ 0 \le y_{ij} \ , \ w_i^+(\omega) \ge 0 \ , w_i^-(\omega) \ge 0 \ , \quad i = 1, \ldots, m \ , \ j = 1, \ldots, n \ . \tag{4.17}$$

This model is a location extension of the transportation model of Williams [1963]. The objective function contains the investment costs for opening plants, the expected production and distribution costs, the expected penalties for extra or insufficient demands, and the expected revenue. This last term is constant because it is assumed that all demands must be satisfied by either direct delivery or some other means reflected in the penalty for unmet demand. The problem only makes sense if q_i^+ is large enough, for example, larger than $E_{\boldsymbol{\xi}}(v_j + t_{ij})$ for all j, although weaker conditions may sometimes suffice. Constraint (4.15) guarantees that distribution only occurs from open plants, i.e., plants such that $x_j = 1$. The constant M represents the maximum possible size of a plant.

Observe that here the variables y_{ij} are first-stage variables. Also observe that in the second stage, the constraints (4.16), (4.17) have a very simple form, as $w_i^+(\omega) = \mathbf{d}_i - \sum_{j=1}^n y_{ij}$ if this quantity is non-negative and $w_i^-(\omega) = \sum_{j=1}^n y_{ij} - \mathbf{d}_i$ otherwise. This is an example of a *second stage with simple recourse*.

Also note that in Cases a and b, the size or capacity of plant j is simply obtained as the sum of the quantity transported from j, namely, $\sum_{i=1}^m d_i y_{ij}$ in Case a and $\sum_{i=1}^m y_{ij}$ in Case b.

c. Uncertain demand, variable distribution pattern

We now consider the case where the distribution pattern can be adjusted to the realization of the random event. This might be the case when uncertainty corresponds to long-term scenarios, of which only one is realized. Then the distribution pattern can be adapted to this particular realization. This also implies that the sizes of the plants cannot be defined as the sum of the quantity distributed, because those quantities depend on the random event. We thus define as before:

$$x_j = \begin{cases} 1 & \text{if plant } j \text{ is open,} \\ 0 & \text{otherwise.} \end{cases}$$

We now let y_{ij} depend on ω with $y_{ij}(\omega) =$ fraction of demand $d_i(\omega)$ served from j and define new variables $w_j =$ size (capacity) of plant j, with unit investment cost g_j.

The model now reads

$$\max \quad -\sum_{j=1}^n c_j x_j - \sum_{j=1}^n g_j w_j + E_{\boldsymbol{\xi}} \max \sum_{i=1}^m \sum_{j=1}^n q_{ij}(\omega) y_{ij}(\omega) \tag{4.18}$$

$$\text{s. t.} \quad x_j \in \{0,1\}\,,\ w_j \geq 0\,, \qquad j = 1,\dots,n\,, \tag{4.19}$$

$$\sum_{j=1}^n y_{ij}(\omega) \leq 1\,, \qquad i = 1,\dots,m\,, \tag{4.20}$$

$$\sum_{i=1}^m d_i(\omega) y_{ij}(\omega) \leq w_j\,, \qquad j = 1,\dots,n\,, \tag{4.21}$$

$$0 \le y_{ij}(\omega) \le x_j \,, \qquad i = 1,\ldots,m \,,\; j = 1,\ldots,n \,, \tag{4.22}$$

where $q_{ij}(\omega) = (r_i - v_j - t_{ij})d_i(\omega)$ now includes the demand $d_i(\omega)$.

Constraint (4.20) indicates that no more than 100% of i 's demand can be served, but that the possibility exists that not all demand is served. Constraint (4.21) imposes that the quantity distributed from plant j does not exceed the capacity w_j decided in the first stage. For the sake of clarity, one could impose a constraint $w_j \le Mx_j$, but this is implied by (4.21) and (4.22). For a discussion of algorithmic solutions of this problem, see Louveaux and Peeters [1992].

d. Stages versus periods; Two-stage versus multistage

In this section, we highlight again the difference in a stochastic program between *stages* and *periods* of times. Consider the case of a distribution firm that makes its plans for the next 36 months. It may formulate a model such as (4.18)–(4.22). The location of warehouses would be first-stage decisions, while the distribution problem would be second-stage decisions. The duration of the first stage would be something like six months (depending on the type of warehouse) and the second stage would run over the 30 remaining months. Although we may think of a problem over 36 periods, a two-stage model is totally relevant. In this case, the only moment where the number of periods is important is when the precise values of the objective coefficients are computed.

In this example, a multistage model becomes necessary if the distribution firm foresees additional periods where it is ready to change the location of the warehouses. In this example, suppose the firm decides that the opening of new warehouses can be decided after one year. A three-stage model can be constructed. The first stage would consist of decisions on warehouses to be built now. The second stage would consist of the distribution patterns between months 7 and 18 as well and new openings decided in month 12 . The third stage would consist of distribution patterns between months 19 and 36 .

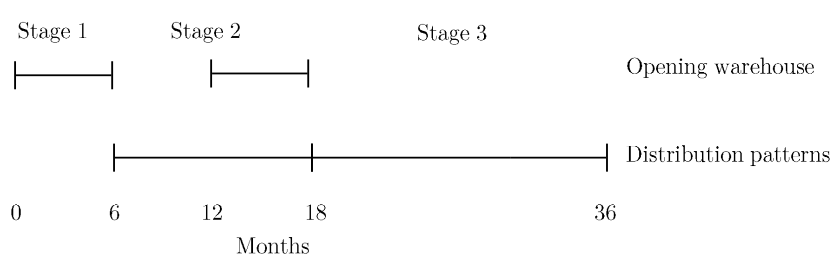

Fig. 2 Three-stage model decisions and times.

Let x^1 and $x^2(\omega_2)$ be the binary vectors representing opening warehouses in stages 1 and 2, respectively. Let $y^2(\omega_2)$ and $y^3(\omega_3)$ be the vectors representing the distribution decisions in stages 2 and 3, respectively, where ω_2 and ω_3 are the states of the world in stages 2 and 3. Assuming each warehouse can only have a fixed size M, the following model can be built:

$$\max \quad -\sum_{j=1}^{n} c_j x_j^1 + \mathrm{E}_{\xi_2} \max\{\sum_{i=1}^{m}\sum_{j=1}^{n} q_{ij}^2(\omega_2) y_{ij}^2(\omega_2) - \sum_{j=1}^{n} c_j^2(\omega_2) x_j^2(\omega_2)$$

$$+ \mathrm{E}_{\xi_3|\xi_2} \max[\sum_{i=1}^{m}\sum_{j=1}^{n} q_{ij}^3(\omega_3) y_{ij}^3(\omega_3)]\}$$

$$\begin{aligned}
\text{s. t.} \quad & \sum_{j=1}^{n} y_{ij}^2(\omega_2) \le 1 , && i = 1,\dots,m , \\
& \sum_{i=1}^{m} d_i(\omega_2) y_{ij}^2(\omega_2) \le M x_j^1 , && j = 1,\dots,n , \\
& \sum_{j=1}^{n} y_{ij}^3(\omega_3) \le 1 , && i = 1,\dots,m , \\
& \sum_{i=1}^{m} d_i(\omega_3) y_{ij}^3(\omega_3) \le M(x_j^1 + x_j^2(\omega_2)) , && j = 1,\dots,n , \\
& x_j^1 + x_j^2(\omega_2) \le 1 , && j = 1,\dots,n , \\
& x_j^1, x_j^2(\omega_2) \in \{0,1\} , && j = 1,\dots,n , \\
& y_{ij}^2(\omega_2), y_{ij}^3(\omega_3) \ge 0 , && i = 1,\dots,m , \; j = 1,\dots,n .
\end{aligned}$$

Multistage programs will be further studied in Section 3.4.

2.5 Random Variables and Risk Aversion

In our view, one can often classify random events and random variables in two major categories. In the first category, we would place uncertainties that recur frequently on a short-term basis. As an example, uncertainty may correspond to daily or weekly demands. This normally leads to a model similar to the one in Section 2.4, Case b (4.b), where allocation cannot be adjusted every time period. It follows that the expectation in the second stage somehow represents a mean over possible values of the random variables, of which many will occur. Thus, the expectation takes into account realizations that might not occur and many realizations that will occur. To fix ideas here, if in Model 4.b the units in the objective function are in a yearly basis and the randomness involves daily or weekly demands, one may expect that the value of the objective of stochastic model will closely match the realized total yearly revenue.

As one interesting example of a real-world application of a location model of this first category, we may recommend the paper by Psaraftis, Tharakan, and Ceder [1986]. It deals with the optimal location and size of equipment to fight oil spills. Occurrence and sizes of spills are random. The sizes of the spills are represented by a discrete random variable taking three possible values, corresponding to small, medium, or large spills. Sadly enough, spills are sufficiently frequent that the expectation may be considered close enough to the mean cost, as just described. Occurrence of spills at a given site is also random. It is described by a Poisson process. By making the assumption of non-concomitant occurrence of spills, all equipment is made available for each spill, which simplifies the second-stage descriptions compared to (4.14)–(4.17).

As a common example, consider revenue management decisions such as those considered in Problem 1.1 for an airline that must determine reservation controls for hundreds of daily flights. This area has become one of the most widespread applications of analytical methods to determining optimal choices under uncertain conditions (see Talluri and van Ryzin [2005]). Airlines routinely solve thousands of these stochastic programs each month and can reasonably expect to receive close to the expected revenue from their decisions each month (if not each day). Risk aversion has little affect in that case.

In the second category, we would place uncertainties that can be represented as scenarios, of which basically only one or a small number are realized. An example in a similar situation to the airline might be the problem of the organizers of the World Cup championship soccer game, which only occurs once every four years, to choose prices and seat allocations to maximize revenues but also to protect against possible losses. This consideration would also be the case in long-term models where scenarios represent the general trend or path of the variables. As already indicated, this is the spirit in which Model 4.c is built. In the second stage, among all scenarios over which expectation is taken, only one is realized. The objective function with only expected values may then be considered a poor representation of risk aversion, which is typically assumed in decision making (if we exclude gambling).

Starting from the von Neumann and Morgenstern [1944] theory of utility, this field of modeling preferences has been developed by economics. Models such as the mean-variance approach of Markowitz [1959] have been widely used. Other methods have been proposed based on mixes of mean-variance and other approaches (see, e.g, Ben-Tal and Teboulle [1986]). From a theoretical point of view, considering a nonlinear utility function transforms the problems into stochastic nonlinear programs, which can require more computational effort than linear versions. In practice, risk aversion is often captured with a piecewise-linear representation, as in the financial planning example in Section 1.2, to maintain a linear problem structure.

One interesting alternative to nonlinear utility models is to include risk aversion in a linear utility model under the form of a linear constraint, called *downside risk* (Eppen, Martin, and Schrage [1989]). The problem there is to determine the type and level of production capacity at each of several locations. Plants produce various types of cars and may be open, closed, or retooled. The demand for each type of car in the medium term is random. The decisions about the locations and configurations of plants have to be made before the actual demands are known.

Scenarios are based on pessimistic, neutral, or optimistic realizations of demands. A scenario consists of a sequence of realizations for the next five years. The stochastic model maximizes the present value of expected discounted cash flows. The linear constraint on risk is as follows: the downside risk of a given scenario is the amount by which profit falls below some given target value. It is thus zero for larger profits. The expected downside risk is simply the expectation of the downside risk over all scenarios. The constraint is thus that the expected downside risk must fall below some level.

To give an idea of how this works, consider a two-stage model similar to (4.1)–(4.4) but in terms of profit maximization, by

$$\max z = c^T x + \mathrm{E}_{\boldsymbol{\xi}}[\max q^T(\omega) y(\omega)]$$

s.t. (4.2)–(4.4).

Then define the target level g on profit. The downside risk $u(\xi)$ is thus defined by two constraints:

$$u(\xi(\omega)) \geq g - q^T(\omega) y(\omega) \tag{5.1}$$

$$u(\xi(\omega)) \geq 0 \,. \tag{5.2}$$

The constraint on expected downside risk is

$$\mathrm{E}_{\boldsymbol{\xi}} u(\boldsymbol{\xi}) \leq l \,, \tag{5.3}$$

where l is some given level. For a problem with a discrete random vector $\boldsymbol{\xi}$, constraint (5.3) is linear. Observe that (5.3) is in fact a first-stage constraint as it runs over all scenarios. It can be used directly in the extensive form. It can also be used indirectly in a sequential manner, by imposing such a constraint only when needed. This can be done in a way similar to the induced constraints for feasibility that we will study in Chapter 5.

2.6 Implicit Representation of the Second Stage

This book is mainly concerned with stochastic programs of the form (4.1)–(4.4), assuming that an adequate and computationally tractable representation of the recourse problem exists. This is not always the case. Two possibilities then exist that still permit some treatment of the problem:

- A closed form expression is available for the expected value function $\mathcal{Q}(x)$.
- For a given first-stage decision x, the expected value function $\mathcal{Q}(x)$ is computable.

These possibilities are described in the following sections.

a. A closed form expression is available for $\mathscr{Q}(x)$

We may illustrate this case by the *stochastic queue median* model (SQM) first proposed by Berman, Larson, and Chiu [1985] from which we take the following in a simplified form. The problem consists of locating an emergency unit (such as an ambulance). When a call arrives, there is a certain probability that the ambulance is already busy handling an earlier demand for ambulance service. In that event, the new service demand is either referred to a backup ambulance service or entered into a queue of other waiting "customers." Here, the first-stage decision consists of finding a location for the ambulance. The second stage consists of the day-to-day response of the system to the random demands. Assuming a first-in, first-out decision rule, decisions in the second stage are somehow automatic. On the other hand, the quality of response, measured, e.g., by the expected service time, depends on the first-stage decision. Indeed, when responding to a call, an ambulance typically goes to the scene and returns to the home location before responding to the next call. The time when it is unavailable for another call is clearly a function of the home location.

Let λ be the total demand rate, $\lambda \geq 0$. Let p_i be the probability that a demand originates from demand region i , with $\sum_{i=1}^{m} p_i = 1$. Let also $t(i,x)$ denote the travel time between location x and call i . On-scene service time is omitted for simplicity. Given facility location x , the expected response time is the sum of the mean-in-queue delay $w(x)$ and the expected travel time $\bar{t}(x)$,

$$\mathscr{Q}(x) = w(x) + \bar{t}(x) \ , \tag{6.1}$$

where

$$w(x) = \begin{cases} \frac{\lambda \bar{t}^{(2)}(x)}{2(1-\lambda \bar{t}(x))} & \text{if } \lambda \bar{t}(x) < 1 \ , \\ 0 & \text{otherwise,} \end{cases} \tag{6.2}$$

$$\bar{t}(x) = \sum_{i=1}^{m} p_i t(i,x) \ , \tag{6.3}$$

and

$$\bar{t}^{(2)}(x) = \sum_{i=1}^{m} p_i t^2(i,x) \ . \tag{6.4}$$

The global problem is then of the form:

$$\min_{x \in X} \mathscr{Q}(x) \ , \tag{6.5}$$

where the first-stage objective function is usually taken equal to zero and X represents the set of possible locations, which typically consists of a network.

It should be clear that no possibility exists to adequately describe the exact sequence of decisions and events in the so-called second stage and that the expected recourse $\mathscr{Q}(x)$ represents the result of a computation assuming the system is in steady state.

b. For a given x, $\mathscr{Q}(x)$ is computable

The deterministic traveling salesperson problem (TSP) consists of finding a Hamiltonian tour of least cost or distance. Following a Hamiltonian tour means that the traveling salesperson starts from her home location, visits all customers, (say $i = 1,\ldots,m$) exactly, and returns to the home location.

Now, assume each customer has a probability p_i of being present. A full optimization that would allow the salesperson to decide the next customer to visit at each step would be a difficult multistage stochastic program. A simpler two-stage model, known as *a priori optimization* is as follows: in the first-stage, an a priori Hamiltonian tour is designed. In the second stage, the a priori tour is followed by skipping the absent customers. The problem is to find the tour with minimal expected cost (Jaillet [1988]).

The exact representation of such a second-stage recourse problem as a mathematical program with binary decision variables might be possible in theory but would be so cumbersome that it would be of no practical value. On the other hand, the expected length of the tour (and thus $\mathscr{Q}(x)$) is easily computed when the tour (x) is given.

Let c_{ij} be the distance between i and j. Assume for simplicity of notation that the given tour is $\{0,1,2,\ldots,n,0\}$ where 0 is the depot.

Define $t(k)$ as the expected length from k till the depot if k is present. Thus we search for $Q(x) = t(0)$.

Start with $t(n+1) = 0$ and $t(n) = c_{n0}$. Let $p_0 = 1$ and $c_{in+1} = c_{i0}$. Then

$$t(k) = \sum_{r=0}^{n-k} \prod_{j=1}^{r} (1 - p_{k+j})\, p_{k+r+1}(c_{kk+r+1} + t(k+r+1)) \quad \text{for } k = n-1,\ldots,0,$$

where the condensed product is equal to 1 if $r = 0$.

This calculation is a backward recursion: assuming k is present, it considers the next present customer to be $k+r+1$ (and thus $k+1$ to $k+r$ being absent) for all possible successors ($k+1$ to $n+1 := 0$).

2.7 Probabilistic Programming

In probabilistic programming, some of the constraints or the objective are expressed in terms of probabilistic statements about first-stage decisions. The description of second-stage or recourse actions is thus avoided. This is particularly useful when the cost and benefits of second-stage decisions are difficult to assess.

For some probabilistic constraints, it is possible to derive a deterministic linear equivalent. A first example was given in Section 1.3. We now detail two other examples where a deterministic linear equivalent is obtained and one where it is not.

a. Deterministic linear equivalent: a direct case

Consider Exercise 1.6.1. An airline wishes to partition a plane of 200 seats into three categories: first, business, economy. Now, assume the airline wishes a special guarantee for its clients enrolled in its loyalty program. In particular, it wants 98% probability to cover the demand of first-class seats and 95% probability to cover the demand of business class seats (by clients of the loyalty program). First-class passengers are covered if they get a first-class seat. Business class passengers are covered if they get either a business or a first-class seat (upgrade). Assume weekday demands of loyalty-program passengers are normally distributed, say $\xi_F \sim N(16,16)$ and $\xi_B \sim N(30,48)$ for first-class and business, respectively. Also assume that the demands for first-class and business class seats are independent.

Let x_1 be the number of first-class seats and x_2 the number of business seats. The probabilistic constraints are simply

$$\mathrm{P}(x_1 \geq \xi_F) \geq 0.98, \tag{7.1}$$

$$\mathrm{P}(x_1 + x_2 \geq \xi_F + \xi_B) \geq 0.95 \,. \tag{7.2}$$

Given the assumptions on the random variables, these probabilistic constraints can be transformed into a deterministic linear equivalent.

Constraint (7.1) can be written as $F_F(x_1) \geq 0.98$, where $F_F(\cdot)$ denotes the cumulative distribution of ξ_F. Now, the 0.98 quantile of the normal distribution is 2.054. As $\xi_F \sim N(16,16)$, $F_F(x_1) \geq 0.98$ is the same as $(x_1 - 16)/4 \geq 2.054$ or $x_1 \geq 24.216$. Thus, the probabilistic constraint (7.1) is equivalent to a simple bound.

Similarly, constraint (7.2) can be written as $F_{FB}(x_1 + x_2) \geq 0.95$, where $F_{FB}(\cdot)$ denotes the cumulative distribution of $\xi_F + \xi_B$. By the independence assumption and the properties of the normal distribution, $\xi_F + \xi_B \sim N(46,64)$. The 0.95 quantile of the standard normal distribution is 1.645. Thus, $F_{FB}(x_1 + x_2) \geq 0.95$ is the same as $(x_1 + x_2 - 46)/8 \geq 1.645$ or $x_1 + x_2 \geq 59.16$.

Thus, the probabilistic constraint (7.2) is equivalent to a linear constraint. We say that (7.2) has a linear deterministic equivalent. This is the desired situation with probabilistic constraints.

b. Deterministic linear equivalent: an indirect case

We now provide an example where finding the deterministic equivalent requires some transformation.

Consider the following covering location problem. Let $j = 1,\dots,n$ be the potential locations with, as usual, $x_j = 1$ if site j is open and 0 otherwise, and c_j the investment cost. Let $i = 1,\dots,m$ be the clients. Client i is served if there exists an open site within distance t_i. The distance between i and j is t_{ij}. Define $N_i = \{j \mid t_{ij} < t_i\}$ as the set of eligible sites for client i. The deterministic covering problem is

$$\min \sum_{j=1}^{n} c_j x_j \tag{7.3}$$

$$\text{s. t.} \sum_{j \in N_i} x_j \geq 1 , \qquad i = 1,\dots,m , \tag{7.4}$$

$$x_j \in \{0,1\} , \qquad j = 1,\dots,n . \tag{7.5}$$

Taking again the case of an ambulance service, one site may cover more than one region or demand area. When a call is placed, the emergency units may be busy serving another call. Let q be the probability that no emergency unit is available at site j. For simplicity, assume this probability is the same for every site (see Toregas et al. [1971]). Then, the deterministic covering constraint (7.4) may be replaced by the requirement that P (at least one emergency unit from an open eligible site is available) $\geq \alpha$ where α is some confidence level, typically 90 or 95%. Here, the probability that none of the eligible sites has an available emergency unit is q to the power $\sum_{j \in N_i} x_j$, so that the probabilistic constraint is

$$1 - q^{\sum_{j \in N_i} x_j} \geq \alpha , \qquad i = 1,\dots,m \tag{7.6}$$

or

$$q^{\sum_{j \in N_i} x_j} \leq 1 - \alpha .$$

Taking the logarithm on both sides, one obtains

$$\sum_{j \in N_i} x_j \geq b \tag{7.7}$$

with

$$b = \left\lceil \frac{\ln(1-\alpha)}{\ln q} \right\rceil , \tag{7.8}$$

where $\lceil a \rceil$ denotes the smallest integer greater than or equal to a. Thus, the probabilistic constraint (7.6) has a linear deterministic equivalent (7.7).

c. Deterministic nonlinear equivalent: the case of random constraint coefficients

The diet problem is a classical example of linear programming (discussed in Dantzig [1963] for the case in Stigler [1945]) . It consists of selecting a number of foods in order to get the cheapest menus that meet the daily requirements in the main nutrients (energy, protein, vitamins,. . .). Consider the data in the introductory example of Chvátal (1980). Polly wants to choose among six foods (oatmeal, chicken, eggs, whole milk, cherry pie and pork with beans). Each food has a given serving size; for instance, a serving of eggs is two large eggs and a serving of pork with beans is 260 grams. Each food has therefore a known content of nutrients. If we take the case of protein, the content is 4 , 32 , 13 , 8 , 4 and 14 grams (grams) of proteins, respectively, for the given serving sizes.

Let $x_1,\ldots,x_6$ represent the number of servings of each product per day. As Polly is a girl of 18 years of age, she needs 55 grams of protein per day. The protein constraint reads as follows:

$$4x_1 + 32x_2 + 13x_3 + 8x_4 + 4x_5 + 14x_6 \geq 55 \ .$$

(We omit here the other constraints and the objective function, which are very important to Polly but not central to our discussion.)

The same book later on contains an interesting discussion on the difficulty to get precise reliable RDA (recommended daily allowances) as well as precise nutrient contents per serving (Chvátal [Chapter 11, pp. 182–187]). Let us concentrate on this second aspect. It is indeed very unlikely that every large egg has exactly 6.5 grams of protein, or every serving of 260 grams of pork with beans has exactly 14 grams of protein. This implies that the nutrient content of each serving is in fact a random variable. Let $a_1,\ldots,a_6$ be the random content in proteins for the six products. The probabilistic constraint reads as follows:

$$\mathrm{P}\left(a_1\, x_1 + a_2\, x_2 + a_3\, x_3 + a_4\, x_4 + a_5\, x_5 + a_6\, x_6 \geq 55\right) \geq \alpha \ . \tag{7.9}$$

Let us now assume the contents of the products are normally distributed, say $a_i \sim N(\mu_i,\sigma_i^2)$, $i = 1,\ldots,6$. We can clearly assume independence between the six products. Then $a_1\, x_1 + a_2\, x_2 + a_3\, x_3 + a_4\, x_4 + a_5\, x_5 + a_6\, x_6 \sim N(\mu,\sigma^2)$ with $\mu = \mu_1\, x_1 + \mu_2\, x_2 + \mu_3\, x_3 + \mu_4\, x_4 + \mu_5\, x_5 + \mu_6\, x_6$ and $\sigma^2 = \sigma_1^2\, x_1^2 + \sigma_2^2\, x_2^2 + \sigma_3^2\, x_3^2 + \sigma_4^2\, x_4^2 + \sigma_5^2\, x_5 + \sigma_6^2\, x_6$.

Classical probabilistic analysis of the normal distribution implies that (7.9) is equivalent to

$$(55-\mu)/\sigma \leq z_{1-\alpha}$$

with $z_{1-\alpha}$ the $(1-\alpha)$ -quantile of the normal distribution. Taking $\alpha = 0.98$, the constraint reads $(55-\mu)/\sigma \leq -2.054$ or $\mu \geq 55 + 2.054 \cdot \sigma$. As $\sigma^2 = \sigma_1^2\, x_1^2 + \sigma_2^2\, x_2^2 + \sigma_3^2\, x_3^2 + \sigma_4^2\, x_4^2 + \sigma_5^2\, x_5^2 + \sigma_6^2\, x_6^2$, this constraint is non-linear and convex.

2.8 Modeling Exercise

In this section, we propose a modeling exercise and comment on a number of possible answers.

a. Presentation

Consider a production or assembly problem. It consists of producing two products, say A and B. They are obtained by assembling two components, say $C1$ and $C2$, in fixed quantities. The following table shows the components usage for the two products:

Components usage	A	B
$C1$	6	10
$C2$	8	5

Components are produced within the plant. Material (and / or operating) costs for $C1$ and $C2$ are 0.4 and 1.2, respectively. The level of production, or capacity, is related to the work-force and the equipment. Each unit of capacity costs 150 and 180 and can produce batches of 60 and 90 components, respectively for $C1$ and $C2$. Current capacity level is $(40, 20)$ batches and cannot be decreased. The total number of batches must not exceed 120. An integer number of batches is not requested here.

In the deterministic case, the demands and unit selling prices are certain and are as follows:

	A	B
Demand	500	200
Unit selling price	50	60

Unmet demand results in lost sales. This does does not imply any additional penalty.

1. Select adequate units for each data. Formulate and solve the deterministic problem.

Then, consider a number of stochastic variants. For the sake of comparison, in all cases, the random variables have expectations which are the corresponding deterministic values.

2. Stochastic prices (known demand).

The selling prices of A and B are described by a random vector, say $\zeta^T = (\zeta_1, \zeta_2)$. The rest of the data is unchanged. Formulate a recourse model in the following cases:

(a) ζ^T takes on the values $(54,56)$, $(50,60)$, and $(46,64)$ with probability 0.3, 0.4 and 0.3 respectively.
(b) ζ_1 takes on the values $(46,50,54)$ with probability 0.3, 0.4 and 0.3; ζ_2 takes on the values $(56,60,64)$ with probability 0.3, 0.4 and 0.3; ζ_1 and ζ_2 are independent.
(c) ζ_1 has a continuous uniform distribution in the range $[46,54]$; ζ_2 has a continuous uniform distribution in the range $[56,64]$; ζ_1 and ζ_2 are independent.
(d) ζ^T takes on the values $(70,50)$, $(50,60)$, $(30,70)$ with probability 0.3, 0.4 and 0.3.
(e) ζ_1 takes on the values $(30,50,70)$ with probability 0.3, 0.4 and 0.3; ζ_2 takes on the values $(50,60,70)$ with probability 0.3, 0.4 and 0.3; ζ_1 and ζ_2 are independent.

3. Stochastic demands (known prices).

The demand levels of A and B are described by a random vector, say $\eta^T = (\eta_1, \eta_2)$. The rest of the data is as in the deterministic model.

(a) Formulate and solve a recourse model when η^T takes on the values $(400,100)$, $(500,200)$, $(600,300)$ with probability 0.3, 0.4 and 0.3.
(b) Assume η_1 and η_2 are independent random variables with normal distributions, $\eta_1 \sim N(500,6000)$ and $\eta_2 \sim N(200,12000)$. Find the optimal solution of the recourse problem if the production of A and B is decided in the first-stage and there is no restriction at all on the number of batches of $C1$ and $C2$.
(c) Consider case (b). Add the restriction that the total number of batches must not exceed 120. Also ensure that the probability that the demand of B is covered must be larger than 80%.

4. Stochastic prices and demands.

Demands and prices are described by three scenarios $S1$, $S2$ and $S3$, as follows.

Demand level	$S1$	$S2$	$S3$
A	700	500	300
B	100	200	300
Unit selling price			
A	45	50	55
B	70	60	50

Formulate and solve a recourse model assuming the three scenarios have probability 0.3, 0.4 and 0.3 respectively.

5. Obtain *EVPI* and *VSS* for some relevant cases among these alternatives.

b. Discussion of solutions

1. Choice of units and deterministic model.

Units are as follows. First, define the unit of time. We may assume here data are given per day for example. Then, demand is the number of units of A and B per day. Selling prices are given as \$ per unit of A and B. The level of production is given by the number of batches (of 60 $C1$ and 90 $C2$) per day. Capacity cost must include work-force cost, operating costs, and depreciation per day. Material costs are \$ per component. The distinction among these costs is important for the stochastic model.

There is more than one formulation for the deterministic problem. The following formulation (M1) is useful in view of later stochastic models. Let

- $x_1 =$ number of batches of $C1$ available for production;
- $x_2 =$ number of batches of $C2$ available for production;
- $x_3 =$ number of units of A produced and sold per day;
- $x_4 =$ number of units of B produced and sold per day.

For batches of $C1$ and $C2$, the objective contains the daily capacity cost. For products A and B, it contains the selling price minus the material costs. (Each unit of A, e.g. has a selling price of \$50. It requests 6 units of C_1 and 8 units of C_2 for a total material cost of \$12. The difference is the objective coefficient 38.) The first two constraints state that the usage of components is smaller than the availability. The third constraint is the upper limit on the number of batches. Demand and capacity bounds follow.

$$
\begin{array}{lll}
\text{(M1)} & z = & \max -150x_1 - 180x_2 + 38x_3 + 50x_4 \\
 & \text{s. t.} & 6x_3 + 10x_4 \le 60x_1, \\
 & & 8x_3 + 5x_4 \le 90x_2, \\
 & & x_1 + x_2 \le 120, \\
 & & 40 \le x_1,\ 20 \le x_2,\ 0 \le x_3 \le 500,\ 0 \le x_4 \le 200.
\end{array}
$$

The optimal solution of (M1) is $z = 5800$, $x_1 = 220/3$, $x_2 = 140/3$, $x_3 = 400$, $x_4 = 200$. Product B is at the maximum corresponding to its demand. All 120 batches of capacity are used. The rest of the solutions follow.

A shorter formulation (M2) is to define two variables:

- $x_1 =$ number of units of A produced and sold per day;
- $x_2 =$ number of units of B produced and sold per day.

This formulation requires computing the margins of A and B. Each unit of A obtains the selling price of \$50. It requires 6 components $C1$ and 8 components $C2$ for a total material cost of \$12. It also requires 6/60 batches of capacity for $C1$ and 8/90 batches for $C2$ at a cost of \$31. The net margin for A is thus \$7 per unit. Similarly, the net margin for B is \$15 per unit. Note that this calculation of the margins of A and B is only valid if there is no unused capacity or unsold product, which is not always the case in a stochastic model. The first two constraints correspond to maintaining at least the existing capacity levels of 40 and 20 respectively. The third constraint corresponds to a maximal capacity level of 120 (each unit of A requires $6/60$ of $C1$ and $8/90$ of $C2$, or $17/90$ capacity units; each unit of B requires $10/60$ of $C1$ and $5/90$ of $C2$ or $20/90$ capacity units). The model also includes the demand constraints and reads as follows:

$$
\begin{aligned}
\text{(M2)} \quad z = \; & \max 7x_1 + 15x_2 \\
\text{s. t.} \quad & 6x_1 + 10x_2 \geq 2400, \\
& 8x_1 + 5x_2 \geq 1800, \\
& 17x_1 + 20x_2 \leq 10800, \\
& 0 \leq x_1 \leq 500 \,,\; 0 \leq x_2 \leq 200.
\end{aligned}
$$

This model has the same optimal solution, $z = 5800$, $x_1 = 400$, $x_2 = 200$, as previously. It is clear in (M2) that the margin of B is larger than that of A. Thus, product B is at the maximum corresponding to its demand. Product A is then reduced from the limit of 120 batches of capacity. The number of batches for $C1$ and $C2$ can be computed from the production of A and B, and are equal to $220/3$ and $140/3$, respectively.

2. Stochastic prices.

The essential modeling question concerns the timing of the decisions. Typically, the capacity decisions are made in the long run. They are first-stage decisions. Sales occur when the price is known. They are always second-stage decisions. Depending on the flexibility of the production process, the decision on the quantity to be produced may be first- or second-stage. We may thus distinguish between two formulations: production is first-stage (M3) or second-stage (M4).

2.1. Production is first-stage.

Let

- $x_1 =$ number of batches of $C1$ available for production;
- $x_2 =$ number of batches of $C2$ available for production;
- $x_3 =$ number of units of A produced per day;
- $x_4 =$ number of units of B produced per day;
- $y_1 =$ number of units of A sold per day;
- $y_2 =$ number of units of B sold per day;

$$\begin{aligned} z = \max &-150x_1 - 180x_2 - 12x_3 - 10x_4 \\ &+ \mathrm{E}_{\boldsymbol{\xi}}(q_1(\omega)\, y_1(\omega) + q_2(\omega)\, y_2(\omega)) \\ \text{s. t. } & 6x_3 + 10x_4 \le 60x_1, \\ & 8x_3 + 5x_4 \le 90x_2, \\ & x_1 + x_2 \le 120 \\ & y_1(\omega) \le x_3,\ y_2(\omega) \le x_4, \\ & 40 \le x_1,\ 20 \le x_2,\ 0 \le x_3,\ 0 \le x_4, \\ & 0 \le y_1(\omega) \le 500,\ 0 \le y_2(\omega) \le 200, \end{aligned}$$

where $\boldsymbol{\xi}^T(\omega) = (q_1(\omega), q_2(\omega)) = \boldsymbol{\zeta}^T(\omega)$ corresponds to the selling prices.

In practice, it is customary to use a simplified notation where the dependence of y and $\boldsymbol{\xi}$ on ω is not made explicit. This (abuse of) notation is used here.

$$\begin{aligned} \text{(M3)} \quad z = \max &-150x_1 - 180x_2 - 12x_3 - 10x_4 \\ &+ \mathrm{E}_{\boldsymbol{\xi}}(q_1\, y_1 + q_2\, y_2) \\ \text{s. t. } & 6x_3 + 10x_4 \le 60x_1, \\ & 8x_3 + 5x_4 \le 90x_2, \\ & x_1 + x_2 \le 120, \\ & y_1 \le x_3,\ y_2 \le x_4, \\ & 40 \le x_1,\ 20 \le x_2,\ 0 \le x_3,\ 0 \le x_4, \\ & 0 \le y_1 \le 500,\ 0 \le y_2 \le 200, \end{aligned}$$

where $\boldsymbol{\xi}^T = (q_1, q_2) = \boldsymbol{\zeta}^T$.

We now transform (M3) as in Section 2.4a. Assuming q_1 and q_2 are never negative (a much needed assumption for the producer to survive), we obtain

$$\begin{aligned} \text{(M3')} \quad z = \max &-150x_1 - 180x_2 - 12x_3 - 10x_4 \\ &+ \mathrm{E}_{\boldsymbol{\xi}}(q_1 \min\{x_3, 500\} + q_2 \min\{x_4, 200\}) \\ \text{s. t. } & 6x_3 + 10x_4 \le 60x_1, \\ & 8x_3 + 5x_4 \le 90x_2, \\ & x_1 + x_2 \le 120, \\ & 40 \le x_1,\ 20 \le x_2,\ 0 \le x_3,\ 0 \le x_4, \end{aligned}$$

or

$$\begin{aligned} \text{(M3'')} \quad z = \max &-150x_1 - 180x_2 - 12x_3 - 10x_4 \\ &+ \mu_1 \min\{x_3, 500\} + \mu_2 \min\{x_4, 200\} \\ \text{s. t. } & 6x_3 + 10x_4 \le 60x_1 \\ & 8x_3 + 5x_4 \le 90x_2 \\ & x_1 + x_2 \le 120 \\ & 40 \le x_1,\ 20 \le x_2,\ 0 \le x_3,\ 0 \le x_4 \end{aligned}$$

where (μ_1, μ_2) is the expectation of $\boldsymbol{\xi}^T$.

As (μ_1, μ_2) is equal to the deterministic selling prices $(50, 60)$, it is easy to show that (M3") has the same optimal solution as the model (M1). This is true for each of the considered cases (a) to (e). To put it another way, if production is decided in the first-stage, the stochastic model where only the selling prices are random can be replaced by a deterministic model with the random prices replaced by their expectations.

2.2. Production is second-stage

Let x_1 and x_2 be as in (M3) and

- $y_1 =$ number of units of A produced and sold per day;
- $y_2 =$ number of units of B produced and sold per day.

$$
\begin{array}{lll}
\text{(M4)} & z = & \max -150x_1 - 180x_2 + \mathrm{E}_{\boldsymbol{\xi}}(q_1\, y_1 + q_2\, y_2) \\
& \text{s. t.} & x_1 \;+\; x_2 \;\le\; 120, \\
& & 6y_1 + 10y_2 \le 60x_1, \\
& & 8y_1 + \;5y_2 \;\le 90x_2, \\
& & 40 \le x_1 \;,\; 20 \le x_2 \;,\; 0 \le y_1 \le 500 \;,\; 0 \le y_2 \le 200,
\end{array}
$$

where $\boldsymbol{\xi}^T = (q_1, q_2) = \boldsymbol{\zeta}^T - (12, 10)$ corresponds to selling prices minus material costs.

Before using formulation (M4), consider the deterministic formulation (M2). As long as the margin of B is larger than the margin of A and the margin of A remains positive, it is optimal to produce and sell $400\,A$ and $200\,B$. If this holds for all realizations of the selling prices, the same optimal solution is obtained for all realizations of $\boldsymbol{\zeta}$. It is thus the optimal solution of the stochastic model. (This will be elaborated in the comments after Proposition 5 of Chapter 4.) The expected margin is simply $\mathrm{E}_{\boldsymbol{\zeta}}(400\zeta_1 + 200\zeta_2 - 26{,}200)$ where $26{,}200$ is the total of the material and capacity costs for the daily production of $400\,A$ and $200\,B$. As (ζ_1, ζ_2) has expectation $(50, 60)$ as in the deterministic model, the expected margin is again the same as in the deterministic model. This situation occurs in cases (a), (b) and (c) of this exercise: the margin of A is $\zeta_1 - 43$, the margin of B is $\zeta_2 - 45$ and the relation $\zeta_2 - 45 \ge \zeta_1 - 43 \ge 0$ holds.

If at some point, the margin of A becomes negative or exceeds that of B, then (M4) is a truly stochastic model. For cases (d) and (e), there are values of the selling prices where the margin of A exceeds that of B. The stochastic model (M4) has to be solved.

In case (d), $\boldsymbol{\zeta}^T$ takes on the values $(70, 50)$, $(50, 60)$, $(30, 70)$ with probability 0.3, 0.4 and 0.3, respectively. First-stage optimal capacity decisions are $(x_1, x_2) = (69.167, 50.833)$. Second-stage optimal production and sale decisions (x_3, x_4) are $(500, 115)$, $(500, 115)$ and $(358.333, 200)$ for the three possible scenarios. The optimal objective value is $z = 5990$.

In case (e), the two random variables $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ are independent, taking three different values each. Thus, the second-stage must consider 9 realizations. The optimal solution is the same as in the deterministic case: first-stage decisions are $(x_1, x_2) = (73.333, 46.667)$, second-stage decisions are $(x_3, x_4) = (400, 200)$, with objective value $z = 5800$.

3. Stochastic demands.

(a) As in Question 2, the first modeling question is the timing of the decisions. Capacity decisions are made in the long run and are first-stage decisions. Sales occur when price is known and are second-stage. The decisions on the quantities to be produced may be first- or second-stage.

(a.1) Production is first-stage.

If production is first-stage, lost sales occur when demand exceeds production. What happens when production exceeds demand is problem dependent. In some situations, excess production may be held in inventory. This would be the case when the randomness represents day-to-day variations in demand. Then excess production is used later to compensate for possible lost sales. Randomness only results in inventory costs. On the other hand, for products such as perishable goods, production is lost ($C1$ and $C2$ could be flour and eggs, A and B could be bread and pastry, e.g.) and lost sales cannot be compensated. The same is true when the randomness describes a set of scenarios of which only one is realized. The scenarios could represent the uncertainty about the success of a new product. If a product is not successful, extra production is lost. If it is very successful, sales are lost to competitors if the production level is insufficient. Or, alternative actions are needed such as subcontracting or overtime.

We now present a formulation (M5) corresponding to a scenario situation (excess production is lost, lost sales are not compensated). The decision variables are the same as in (M3).

$$
\begin{aligned}
\text{(M5)}\quad z = \; & \max -150x_1 - 180x_2 - 12x_3 - 10x_4 \\
& \qquad + \mathrm{E}_{\boldsymbol{\xi}}(50y_1 + 60y_2) \\
\text{s. t.}\; & 6x_3 + 10x_4 \le 60x_1, \\
& 8x_3 + 5x_4 \le 90x_2, \\
& x_1 + x_2 \le 120, \\
& y_1 \le x_3 \;,\; y_2 \le x_4, \\
& 40 \le x_1 \;,\; 20 \le x_2 \;,\; 0 \le x_3 \;,\; 0 \le x_4 \;, \\
& 0 \le y_1 \le d_1 \;,\; 0 \le y_2 \le d_2,
\end{aligned}
$$

where $\boldsymbol{\xi}^T = (d_1, d_2) = \boldsymbol{\eta}^T$ correspond to the demand level.

The first-stage optimal capacity decisions are $(x_1, x_2) = (56.667, 41.111)$. The second-stage optimal production and sale decisions (x_3, x_4) are $(400, 100)$ in the three possible scenarios. The optimal objective value is $z = 4300$. Observe that the production is set to meet the lowest possible demand.

(a.2) Production is second-stage.

If production is second-stage, lost sales occur when the available production capacities are insufficient to cover the demand. Excess production does not occur as the level of production can be adjusted to the downside. The decision variables are the same as in (M4). Formulation (M6) reads as follows:

$$\begin{array}{lll}
\text{(M6)} & z = & \max -150x_1 - 180x_2 + \mathrm{E}_{\boldsymbol{\xi}}(38y_1 + 50y_2) \\
& \text{s. t.} & x_1 + x_2 \le 120, \\
& & 6y_1 + 10y_2 \le 60x_1, \\
& & 8y_1 + 5y_2 \le 90x_2, \\
& & 40 \le x_1 \ , \ 20 \le x_2 \ , \ 0 \le y_1 \le d_1 \ , \ 0 \le y_2 \le d_2,
\end{array}$$

where $\boldsymbol{\xi}^T = (d_1, d_2) = \boldsymbol{\eta}^T$ corresponds to the demand level.

The first-stage optimal capacity decisions are $(x_1, x_2) = (67.083, 41.111)$. The second-stage optimal production and sale decisions (x_3, x_4) are $(400, 100)$, $(337.5, 200)$ and $(337.5, 200)$ for the three possible scenarios. The optimal objective value is $z = 4575$. Observe that the capacity limit of 120 batches is not fully used.

(b) We consider a variant of formulation (M5) where the only constraints on x_1 and x_2 are the components usage:

$$\begin{array}{lll}
\text{(M7)} & z = & \max -150x_1 - 180x_2 - 12x_3 - 10x_4 \\
& & \quad + \mathrm{E}_{\boldsymbol{\xi}}(50\min\{x_3, d_1\} + 60\min\{x_4, d_2\}) \\
& \text{s. t.} & 6x_3 + 10x_4 \le 60x_1, \\
& & 8x_3 + 5x_4 \le 90x_2, \\
& & 0 \le x_1 \ , \ 0 \le x_2 \ , \ 0 \le x_3 \ , \ 0 \le x_4,
\end{array}$$

where $\boldsymbol{\xi}^T = (d_1, d_2) = \boldsymbol{\eta}^T$ corresponds to the demand level.

Clearly, the two constraints are always tight. Replacing x_1 by $(6x_3 + 10x_4)/60$ and x_2 by $(8x_3 + 5x_4)/90$, the model becomes

$$\begin{aligned}
z = \ & \max\{-43x_3 - 45x_4 + \mathrm{E}_{\boldsymbol{\xi}}(50\min\{x_3, d_1\} \\
& \quad + 60\min\{x_4, d_2\}) \mid 0 \le x_3, \ 0 \le x_4\} \ ,
\end{aligned}$$

or

$$\begin{aligned}
\text{(M7')} \quad z = \ & \max\{-43x_3 + 50\mathrm{E}_{\boldsymbol{\xi}_1}\min\{x_3, \boldsymbol{\xi}_1\} - 45x_4 \\
& \quad + 60\mathrm{E}_{\boldsymbol{\xi}_2}\min\{x_4, \boldsymbol{\xi}_2\} \mid 0 \le x_3 \ , \ 0 \le x_4\}.
\end{aligned}$$

This optimization is separable in x_3 and x_4. Both variables will be nonzero. So, we are searching twice for the unconstrained minimum of an expression of the form $-ax + b\mathscr{Q}(x)$, with $\mathscr{Q}(x) = \mathrm{E}_{\boldsymbol{\xi}}\min\{x, \boldsymbol{\xi}\}$ and $\boldsymbol{\xi} \sim N(\mu, \sigma^2)$. From Exercise 2.8.2, we obtain that $\mathscr{Q}'(x) = 1 - F(x)$. As $\mathscr{Q}''(x) = -f(x)$, the second-order conditions are satisfied. Thus the unconstrained minimum is obtained for $\mathscr{Q}'(x) = a/b$, i.e. $1 - F(x) = a/b$.

Denote by $F_i(\cdot)$ the cumulative distribution of $\boldsymbol{\xi}_i$, $i = 1, 2$. For x_3, the unconstrained optimum satisfies $1 - F_1(x_3) = 43/50$, or $F_1(x_3) = 0.14$. It corresponds to a quantile $q = -1.08$ and a decision $x_3 = 500 - 1.08\sqrt{6000} = 416.34$. For x_4, we have $1 - F_2(x_4) = 45/60$, or $F_2(x_4) = 0.25$. It corresponds to a quartile $q = -0.675$ and a decision $x_4 = 200 - 0.675\sqrt{12000} = 126.06$. For the sake of comparison, we may compute $x_1 = (6x_3 + 10x_4)/60 = 62.644$ and $x_2 = (8x_3 + 5x_4)/90 = 44.011$. Also, using the closed form expression of $\mathscr{Q}(x)$, (see again Exercise 2.8.2), one can obtain the optimal value of z.

(c) Requesting that the probability that the demand of B is covered must be larger than 80% is $\mathrm{P}(x_4 \geq \xi_2) \geq 0.8$ or $F_2(x_4) \geq 0.8$. The 0.8 quantile is 0.84. Thus, $F_2(x_4) \geq 0.8$ is equivalent to $(x_4 - \mu_2)/\sigma_2 \geq 0.8$, or $x_4 \geq 200 + 0.84\sqrt{12000}$, or $x_4 \geq 292.02$.

The model to solve is:

$$\begin{aligned}(\text{M8})\quad z = \max\{&-43x_3 + 50\mathrm{E}_{\xi_1}\min\{x_3,\xi_1\} - 45x_4\\ &+ 60\mathrm{E}_{\xi_2}\min\{x_4,\xi_2\} \mid 0 \leq x_3,\\ &292.02 \leq x_4\ ,\ 17x_3 + 20x_4 \leq 10800\},\end{aligned}$$

where the constraint on the 120 batches has been transformed as in (M2).

By applying the Karush-Kuhn-Tucker conditions (see Review Section 2.11c.), one can show that $(x_3,x_4) = (291.74, 292.02)$ is the optimal solution.

4. Just as in the previous cases, there are two possible formulations as the production decisions may be first- or second-stage. Model (M9) corresponds to first-stage production while (M10) corresponds to second-stage production.

$$\begin{aligned}(\text{M9})\quad z = \ &\max -150x_1 - 180x_2 - 12x_3 - 10x_4\\ &\qquad + \mathrm{E}_{\xi}(q_1\, y_1 + q_2\, y_2)\\ \text{s. t.}\ \ &6x_3 + 10x_4 \leq 60x_1,\\ &8x_3 + 5x_4 \leq 90x_2,\\ &x_1 + x_2 \leq 120,\\ &y_1 \leq x_3\ ,\ y_2 \leq x_4,\\ &40 \leq x_1\ ,\ 20 \leq x_2\ ,\ 0 \leq x_3\ ,\ 0 \leq x_4\ ,\\ &0 \leq y_1 \leq d_1\ ,\ 0 \leq y_2 \leq d_2,\end{aligned}$$

where $\xi^T = (q_1, q_2, d_1, d_2)$, with q_1 and q_2 the selling prices and d_1 and d_2 the demands jointly defined in a scenario. Thus $\xi^T = (45,70,700,100), (50,60,500,200)$ and $(55,50,300,300)$ with probability 0.3, 0.4, and 0.3 respectively. The optimal solution is $z = 3600$, $(x_1,x_2) = (46.667, 32.222)$ with corresponding $(x_3,x_4) = (300,100)$. The second-stage decisions are $(y_1,y_2) = (300,100)$ in all three scenarios. As the production cannot be adapted to the demand, the optimal solution is to plan for the lowest demand and the expected margin is low.

$$\begin{aligned}(\text{M10})\quad z = \ &\max -150x_1 - 180x_2 + \mathrm{E}_{\xi}(q_1\, y_1 + q_2\, y_2)\\ \text{s. t.}\ \ &x_1 + x_2 \leq 120\\ &6y_1 + 10y_2 \leq 60x_1,\\ &8y_1 + 5y_2 \leq 90x_2,\\ &40 \leq x_1\ ,\ 20 \leq x_2\ ,\ 0 \leq y_1 \leq d_1\ ,\ 0 \leq y_2 \leq d_2,\end{aligned}$$

where $\xi^T = (q_1, q_2, d_1, d_2)$ with q_1 and q_2 the selling prices minus the material costs and d_1 and d_2 the demands. Thus, $\xi^T = (33,60,700,100)$, $(38,50,500,200)$ and $(43,40,300,300)$ with probability 0.3, 0.4, and 0.3. The optimal solution is $z = 4048.75$, $(x_1,x_2) = (73.333, 46.667)$. The second-stage decisions are $(y_1,y_2) = (462.5,100)$, $(400,200)$ and $(300,260)$ in the three scenarios. While obtaining the optimal solution of (M10) with your favorite LP solver, you may observe that there is a high shadow price for the maximum number of batches.

Exercises

1. Consider Exercise 1 of Section 1.6.

   (a) Show that this is a two-stage stochastic program with first-stage integer decision variables. Observe that, for a random variable with integer realizations, the second-stage variables can be assumed continuous because the optimal second-stage decisions are automatically integer. Assume that Northam revises its seating policy every year. Is a multistage program needed?

   (b) Assume that the data in Exercise 1 correspond to the demand for seat reservations. Assume that there is a 50% probability that all clients with a reservation effectively show up and that 10 or 20% no-shows occur with equal probability. Model this situation as a three-stage program, with first-stage decisions as before, second-stage decisions corresponding to the number of accepted reservations, and third-stage decisions corresponding to effective seat occupation. Show that the third stage is a simple recourse program with a reward for each occupied seat and a penalty for each denied reservation.

   (c) Consider now the situation where the number of seats has been fixed to 12, 24, and 140 for the first class, business class, and economy class, respectively. Assume the top management estimates the reward of an occupied seat to be 4, 2, and 1 in the first class, business class, and economy class, respectively, and the penalty for a denied reservation is 1.5 times the reward. Model the corresponding problem as a recourse program. Find the optimal acceptance policy with the data of Exercise 1 in Section 1.6 and no-shows as in (b) of the current exercise. To simplify, assume that passengers with a denied reservation are not seated in a higher class even if a seat is available there.

2. Let $\mathscr{Q}(x) = \mathrm{E}_{\boldsymbol{\xi}} \min\{x, \boldsymbol{\xi}\}$.

   (a) Obtain a closed form expression for $\mathscr{Q}(x)$ when $\boldsymbol{\xi}$ follows a Poisson distribution.

   (b) Obtain a closed form expression for $\mathscr{Q}(x)$ when $\boldsymbol{\xi}$ follows a normal distribution. (Hint: for a normal distribution, the relation $\xi f(\xi) = \mu f(\xi) - \sigma^2 f'(\xi)$ holds for any given ξ.)

   (c) Assume $\boldsymbol{\xi}$ has a continuous distribution. Show that $\mathscr{Q}'(x) = 1 - F(x)$.

3. Consider an airplane with x seats. Assume passengers with reservations show up with probability 0.90, independently of each other.

(a) Let $x = 40$. If 42 passengers receive a reservation, what is the probability that at least one is denied a seat.
(b) Let $x = 50$. How many reservations can be accepted under the constraint that the probability of seating all passengers who arrive for the flight is greater than 90% ?

4. Consider the design problem in Section 1.4. Suppose the design decision does not completely specify x in $(1.4.1)$, but the designer only knows that if a value $\hat{x}$ is specified then $x \in [.99\hat{x}, 1.01\hat{x}]$. Suppose a uniform distribution for x is assumed initially on this interval. How would the formulation in Section 1.4 be modified to account for information as new parts are produced?
5. Consider the example in Section 2.7a.

   (a) One may feel uncomfortable with the deterministic linear equivalent yielding a non-integer number of seats. Show how to cope with this.
   (b) One may also feel uncomfortable with the demands represented by normal distributions. Show that deterministic linear equivalents are also obtained if $\boldsymbol{\xi}_F \sim P(3)$ and $\boldsymbol{\xi}_B \sim P(4)$ for example.

2.9 Alternative Characterizations and Robust Formulations

While the main focus of this book is on problems that can be represented in the form in (4.1–4.4) as stochastic linear programs, this formulation can still represent a wide range of risk preferences. As observed in Section 2.5, an expected von Neumann-Morgenstern concave utility objective can be represented as a piecewise-linear function. For example, if the utility function is $U(-q(\omega)^T y(\omega) - \gamma)$ where γ is a scaling parameter for fitting the function, then an additional set of variables $y'(\omega)_j$ with bounds u_j and slopes $-q'_j$ such that $0 \le y'(\omega)_j \le u_j$, $-q'_j \ge -q'_{j+1}$, and for $j = 0, \ldots, J$ can be defined with an additional linear constraint as:

$$-y'_0 + \sum_{j=1} y'_j(\omega) - q(\omega)^T y(\omega) = \gamma, \tag{9.1}$$

and with a new recourse function objective to minimize

$$-q'_0 y_0(\omega) + \sum_{j=1}^{J} q'_j y(\omega). \tag{9.2}$$

The parameters γ, q', and u' can be chosen to fit the utility function U as closely as desired while maintaining the same linear optimization form as in (4.1–4.4).

Other risk–measures may be included in the objective and as fixed or probabilistic constraints. A common use of these constraints in financial applications is to maximize expected return subject to a constraint on *value–at–risk* (*VaR*), the greatest loss in portfolio value that can occur with a given probability α, defined as

$$VaR_{\alpha}(q(\omega)^T y(\omega)) = \min\{t | \mathrm{P}(q(\omega)^T y(\omega) \leq t) \geq \alpha\}. \tag{9.3}$$

A VaR constraint to limit losses to be no greater than $\bar{t}$ with probability at most α can then be written as

$$\mathrm{P}(q(\omega)^T y(\omega) \leq \bar{t}) \geq \alpha, \tag{9.4}$$

since this ensures that $VaR_{\alpha}(q(\omega)^T y(\omega)) \leq \bar{t}$.

A criticism of VaR as a measure of risk is that it does not have the useful property of subadditivity such that the VaR of the sum of two random variables is at most the the sum of the VaR 's of each individual random variable. The subadditive property is part of the set of axioms that define coherent risk measures (see Artzner, Delbaen, Eber, and Heath [1999]), such that $R(\cdot)$ is a *coherent risk measure* if the following hold:

Definition 2.1. 1. *subadditivity*: $R(\xi+\zeta) \leq R(\xi)+R(\zeta)$ for any random variables ξ and ζ;
2. *positive homogeneity (of degree one)*: $R(\lambda\xi) = \lambda R(\xi)$ for all $\lambda \geq 0$;
3. *monotonicity*: $R(\xi) \leq R(\zeta)$ whenever $\xi \preceq \zeta$, where $\preceq$ indicates first-order stochastic dominance,i.e., $\mathrm{P}(\xi \leq t) \geq \mathrm{P}(\zeta \leq t), \forall t$;
4. *translation invariance*: $R(\xi+t) = R(\xi)+t$ for any $t \in \Re$.

A related risk measure to VaR, called the *conditional value-at-risk* ($CVaR$), can be defined to avoid the potential problems of a non-subadditive risk measure by taking the conditional expectations over losses in excess of VaR. For random loss ξ with distribution function P, the α -confidence level is then defined as

$$CVaR_{\alpha}(\xi) = \mathrm{E}_{P_{\alpha}}[\xi], \tag{9.5}$$

where P_{α} is the distribution function defined by

$$P_{\alpha}(t) = \begin{cases} 0 & \text{if } t < VaR_{\alpha}(\xi); \\ \frac{P(t)-\alpha}{1-\alpha} & \text{if } t \geq VaR_{\alpha}(\xi). \end{cases} \tag{9.6}$$

As shown by Rockafellar and Uryasev [2000,2002], $CVaR$ satisfies all of the axioms for a coherent risk measure (Exercise 3) and has a convenient representation as the solution to the following optimization problem:

$$CVaR_{\alpha}(\xi) = \min_t \; t + \frac{1}{1-\alpha}\mathrm{E}_P[(\xi - t)^+], \tag{9.7}$$

which can also be written as the linear program:

$$\min \; t + \frac{1}{1-\alpha}\mathrm{E}_P[y(\omega)] \tag{9.8}$$

$$\text{s. t. } \xi(\omega) - y(\omega) \leq t, \text{ a. s.} \tag{9.9}$$

$$y(\omega) \geq 0, \text{ a. s.} \tag{9.10}$$

With the representation in (9.8), a risk constraint to limit $CVaR_\alpha$ to be less than $\bar{t}$ can be constructed similarly to the probabilistic constraint in (9.4) or the downside risk constraint in (5.3) with additional linear constraints and variables $y'(\omega)$ as follows:

$$t + \frac{1}{1-\alpha}\mathrm{E}[y'(\omega)] \le \bar{t} \tag{9.11}$$

$$-t + q(\omega)^T y(\omega) - y'(\omega) \le 0,\ a.s., \tag{9.12}$$

$$y'(\omega) \ge 0, a.s. \tag{9.13}$$

The use of coherent risk measures has another useful interpretation that R is a coherent risk measure if and only if there is a class of probability measure $\mathscr{P}$ such that $R(\xi)$ equals the highest expectation of ξ with respect to members of this class (see Huber [1981]):

$$R(\xi) = \sup_{P\in\mathscr{P}} \mathrm{E}_P[\xi]. \tag{9.14}$$

This representation provides a worst-case view of the risk, which is discussed in more detail in Chapter 8.

One worst-case version of the approach in (9.14) is to let $\mathscr{P}$ correspond to any distribution with support in a given range or uncertainty set. This worst-case type of risk-measure is called *robust* so that optimization models including a robust risk-measure of this form are *robust optimization* models. A robust version of the two–stage stochastic program can then be written as:

$$\begin{aligned} \min_x \max_{\xi\in\Xi}\ & c^T x + Q(x,\xi) \\ \text{s. t.}\quad & Ax = b, \\ & x \ge 0. \end{aligned} \tag{9.15}$$

Depending on the properties of Ξ, robust optimization models can be tractable linear or conic optimization models. A variety of results in the area appear in Bertsimas and Sim [2006], Ben-Tal and Nemirovski [2002] with multi-period extensions also appearing, for example, in Ben-Tal, Boyd, and Nemirovski [2006] and Bertsimas, Iancu, and Parrilo [2010].

Exercises

1. Give an example of random variables ξ and ζ where $VaR_\alpha(\xi+\zeta) > VaR_\alpha(\xi) + VaR_\alpha(\zeta)$ for some $0 < \alpha < 1$.
2. Show that VaR satisfies the axioms of positive homogeneity, monotonicity, and translation independence.
3. Show that $CVaR$ satisfies all of the axioms for a coherent risk measure.
4. Give a class of probability distribution $\mathscr{P}$ such that $CVaR$ solves (9.14).

5. Find the robust formulation of the two-stage model (9.15) when uncertainty is only in the right-hand side $h \in \Xi = [l,u]$, a rectangular region.
6. Find the robust formulation of the two-stage model (9.15) when uncertainty is only in the right-hand side $h \in \Xi = \{h|(h-\mu)^T V(h-\mu) \leq 1\}$, an ellipsoidal region.

2.10 Relationship to Other Decision-Making Models

The stochastic programming models considered in this section illustrate the general form of a stochastic program. While this form can apply to virtually all decision-making problems with unknown parameters, certain characteristics typify stochastic programs and form the major emphasis of this book. In general, stochastic programs are generalizations of deterministic mathematical programs in which some uncontrollable data are not known with certainty. The key features are typically many decision variables with many potential values, discrete time periods for decisions, the use of expectation functionals for objectives, and known (or partially known) distributions. The relative importance of these features contrasts with similar areas, such as statistical decision theory, decision analysis, dynamic programming, Markov decision processes, and stochastic control. In the following subsections, we consider these other areas of study and highlight the different emphases.

a. Statistical decision theory and decision analysis

Wald [1950] developed much of the foundation of optimal statistical decision theory (see also DeGroot [1970] and Berger [1985]). The basic motivation was to determine best levels of variables that affect the outcome of an experiment. With variables x in some set X , random outcomes, $\omega \in \Omega$, an associated distribution, $F(\omega)$, and a reward or loss associated with the experiment under outcome ω of $r(x,\omega)$, the basic problem is to find $x \in X$ to

$$\max \mathrm{E}_{\omega}[r(x,\omega)|F] = \max \int_{\omega} r(x,\omega) dF(\omega). \tag{10.1}$$

The problem in (10.1) is also the fundamental form of stochastic programming. The major differences in emphases between the fields stem from underlying assumptions about the relative importance of different aspects of the problem.

In stochastic programming, one generally assumes that difficulties in finding the form of the function r and changes in the distribution F as a function of actions are small in comparison to finding the expectations with known distributions and an optimal value x with all other information known. The emphasis is on finding a solution after a suitable problem statement in the form (10.1) has been found.

For example, in the simple farming example in Section 1.1, the number of possible planting configurations (even allowing only whole-acre lots) is enormous. Enumerating the possibilities would be hopeless. Stochastic programming avoids such inefficiencies through an optimization process.

We might suppose that the fields or crop varieties are new and that the farmer has little direct information about yields. In this case, the yield distribution would probably start as some prior belief but would be modified as time went on. This modification and possible effects of varying crop rotations to obtain information are the emphases from statistical decision theory. If we assumed that only limited variation in planting size (such as 50-acre blocks) was possible, then the combinatorial nature of the problem would look less severe. Enumeration might then be possible without any particular optimization process. If enumeration were not possible, the farmer might still update the distributions and objectives and use stochastic programming procedures to determine next year's crops based on the updated information.

In terms of (10.1)), statistical decision theory places a heavy emphasis on changes in F to some updated distribution $\hat{F}_x$ that depends on a partial choice of x and some observations of ω. The implied assumption is that this part of the analysis dominates any solution procedure, as when X is a small finite set that can be enumerated easily.

Decision analysis (see, e.g., Raiffa [1968]) can be viewed as a particular part of optimal statistical decision theory. The key emphases are often on acquiring information about possible outcomes, on evaluating the utility associated with various outcomes, and on defining a limited set of possible actions (usually in the form of a decision tree). For example, consider the capacity expansion problem in Section 1.3. We considered a wide number of alternative technology levels and production decisions. In that model, we assumed that demand in each period was independent of the demand in the previous period. This characteristic gave the block separability property that can allow efficient solutions for large problems.

A decision analytic model might apply to the situation where an electric utility's demand depends greatly on whether a given industry locates in the region. The decision problem might then be broken into separate stochastic programs depending on whether the new industry demand materializes and whether the utility starts on new plants before knowing the industry decision. In this framework, the utility first decides whether to start its own projects. The utility then observes whether the new industry expands into the region and faces the stochastic program form from Section 1.4 with four possible input scenarios about the available capacity when the industry's location decision is known (see Figure 3).

The two stochastic programs given each initial decision allow for the evaluation of expected utility given the two possible outcomes and two possible initial decisions. The actual initial decision taken on current capacity expansion would then be made by taking expectations over these two outcomes.

Separation into distinct possible outcomes and decisions and the realization of different distributions depending on the industry decision give this model a decision analysis framework. In general, a decision analytic approach would probably also consider multiple attributes of the capacity decisions (for example, social costs for a

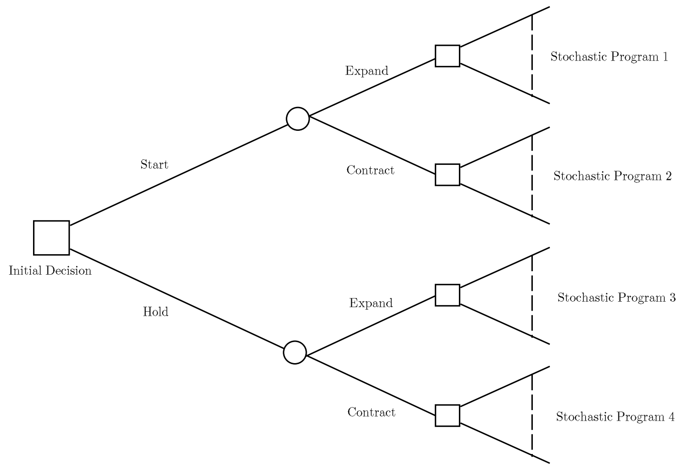

Fig. 3 Decision tree for utility with stochastic programs on leaves.

given location) and would concentrate on the value of risk in the objective. It would probably also entail consideration of methods for obtaining information about the industry's decision and contingent decisions based on the outcomes of these investigations. Of course, these considerations can all be included in a stochastic program, but they are not typically the major components of a stochastic programming analysis.

b. Dynamic programming and Markov decision processes

Much of the literature on stochastic optimization considers dynamic programming and Markov decision processes (see, e.g., Heyman and Sobel [1984], Bellman [1957], Ross [1983], and Kall and Wallace [1994] for a discussion relating to stochastic programming). In these models, one searches for optimal actions to take at generally discrete points in time. The actions are influenced by random outcomes and carry one from some state at some stage t to another state at stage $t+1$. The emphasis in these models is typically in identifying finite (or, at least, low-dimensional) state and action spaces and in assuming some Markovian structure (so that actions and outcomes only depend on the current state).

With this characterization, the typical approach is to form a backward recursion resulting in an optimal decision associated with each state at each stage. With large state spaces, this approach becomes quite computationally cumbersome although it does form the basis of many stochastic programming computation schemes as given in Chapter 6. Another approach is to consider an infinite horizon and use discounting to establish a stationary policy (see Howard [1960] and Blackwell [1965]) so that one need only find an optimal decision associated with a state for any stage.

A typical example of this is in investment. Suppose that instead of saving for a specific time period in the example of Section 1.2, you wish to maximize a discounted expected utility of wealth in all future periods. In this case, the state of the system is the amount of wealth. The decision or action is to determine what amount of the wealth to invest in stock and bonds. We could discretize to varying wealth levels and then form a problem as follows:

$$\max \sum_{t=1}^{\infty} \rho^t \mathrm{E}\,[q\mathbf{y}(t) - r\mathbf{w}(t)] \tag{10.2}$$

$$\begin{aligned} \text{s. t.} \quad & x(1,1) + x(2,1) = b, \\ & \boldsymbol{\xi}(1,t)\mathbf{x}(1,t) + \boldsymbol{\xi}(2,t)\mathbf{x}(2,t) - \mathbf{y}(t) + \mathbf{w}(t) = G, \\ & \boldsymbol{\xi}(1,t)\mathbf{x}(1,t) + \boldsymbol{\xi}(2,t)\mathbf{x}(2,t) = \mathbf{x}(1,t+1) + \mathbf{x}(2,t+1), \\ & \mathbf{x}(i,t), \mathbf{y}(t), \mathbf{w}(t) \geq 0, \quad \mathbf{x} \in \mathcal{N}, \end{aligned}$$

where $\mathcal{N}$ is the space of nonanticipative decisions and ρ is some discount factor. This approach could lead to finding a stationary solution to

$$\begin{aligned} z(b) = \max_{x(1)+x(2)=b} \{ & \mathrm{E}\,[-q(G - \boldsymbol{\xi}(1)x(1) - \boldsymbol{\xi}(2)x(2))^- \\ & -r(G - \boldsymbol{\xi}(1)x(1) - \boldsymbol{\xi}(2)x(2))^+ + \rho \mathrm{E}\,[z(\boldsymbol{\xi}(1)x(1) + \boldsymbol{\xi}(2)x(2))]\}. \end{aligned} \tag{10.3}$$

Again, problem (10.2) fits the general stochastic programming form, but particular solutions as in (10.3) are more typical of Markov decision processes. These are not excluded in stochastic programs, but stochastic programs generally do not include the Markovian assumptions necessary to derive (10.3).

c. Machine learning and online optimization

While Markov decision problems have the general character of stochastic programs of including a distribution over some set of uncertain parameters, online optimization problems involve a changing objective (perhaps chosen adversarially) without knowledge of the choice and only considering the history of observations. The objective is then to choose $x^1, x^2, \ldots$ sequentially to minimize

$$\sum_{t=1}^{H} f^t(x^t), \tag{10.4}$$

where H may increase without bound and each x^t is chosen only with knowledge of $x^1, \ldots, x^{t-1}$ and $f^1(x^1), \ldots, f^{t-1}(x^{t-1})$. Performance is measured in terms of regret, which refers to the difference relative to best possible choices taken *ex post*, i.e.,

$$\text{regret}_H = \sum_{t=1}^{H} f^t(x^t) - \min_{x \in X} \sum_{t=1}^{H} f^t(x), \tag{10.5}$$

where X is some feasible region.

The emphasis in this stream of literature is on algorithms with provable regret bounds. For convex objectives, stochastic search methods (as in Chapter 9) can obtain bounds on regret_H, such as $O(H^{3/4})$, $O(\sqrt{H})$, and $O(\log H)$ depending on properties of f^t and observability of the function (see, respectively, Hazan, Kalai, Kale, and Agarwal [2006], Zinkerich [2003], Flaxman, Kalai, and McMahon [2004]).

d. Optimal stochastic control

Stochastic control models are often similar to stochastic programming models. The differences are mainly due to problem dimension (stochastic programs would generally have higher dimension), emphases on control rules in stochastic control, and more restrictive constraint assumptions in stochastic control. In many cases, the distinction is, however, not at all clear.

As an example, suppose a more general formulation of the financial model in Section 1.2. There, we considered a specific form of the objective function, but we could also use other forms. For example, suppose the objective was generally stated as minimizing some cost $r_t(\mathbf{x}(t), \mathbf{u}(t))$ in each time period t, where $\mathbf{u}(t)$ are the controls $u((i,j),t,s)$ that correspond to actual transactions of exchanging asset i into asset j in period t under scenario s. In this case, problem (1,2.2) becomes:

$$\begin{aligned}
\min z = \sum_s p(s)\Big(\sum_{t=1}^{H} r_t(x(t,s), u(t,s), s)\Big) \\
\text{s. t.} \qquad x(0,s) &= b, \\
x(t,s) + \xi(s)^T u(t,s) &= x(t+1,s), t = 0, \ldots, H, \\
x(s), u(s) \text{ nonanticipative}, &
\end{aligned} \tag{10.6}$$

where $\xi(s)$ represents returns on investments minus transaction costs. Additional constraints may be incorporated into the objective of (10.6) through penalty terms.

Problem (10.6) is fairly typical of a discrete time control problem governed by a linear system. The general emphasis in control approaches to such problems is for *linear, quadratic, Gaussian* (LQG) models (see, for example, Kushner [1971], Fleming and Rishel [1975], and Dempster [1980]), where we have a linear system as earlier, but where the randomness is Gaussian in each period (for example, ξ is known but the state equation for $x(t+1,s)$ includes a Gaussian term), and r_t is quadratic. In these models, one may also have difficulty observing x so that an additional observation variable $y(t)$ may be present.

LQG models can also include forms of risk aversion as, for example, in Whittle [1990]. In this model, instead of an additively time-separable model as generally used here, the objective to minimize becomes:

$$\frac{2}{\theta}\log \mathrm{E}\,[e^{\theta \sum_{t=1}^{H} (\mathbf{x}^t)^T Q^t \mathbf{x}^t + (\mathbf{u}^t)^T R^t \mathbf{u}^t}], \tag{10.7}$$

where $x^{t+1} = A^t x^t + B^t u^t + \varepsilon^t$. A useful property is that this objective avoids some of the issues with time-additive utility functions that do not appear consistent with preferences (as, for example, discussed in Kreps and Porteus [1979], Epstein and Zinn [1989]). A minimizing solution also has a min-max characterization as in robust optimization models and the max-min utility function proposed in Gilboa and Schmeidler [1989] (see Exercise 3 and Hansen and Sargent [1995]).

The LQG problem leads to Kalman filtering solutions (see, for example, Kalman [1969]). Various extensions of this approach are also possible, but the major emphasis remains on developing controls with specific decision rules to link observations directly into estimations of the state and controls. In stochastic programming models, general constraints (such as non-negative state variables) are emphasized. In this case, most simple decision rules forms (such as when u is a linear function of state) fail to obtain satisfactory solutions (see, for example, Gartska and Wets [1974]). For this reason, stochastic programming procedures tend to search for more general solution characteristics.

Stochastic control procedures may, of course, apply but stochastic programming tends to consider more general forms of interperiod relationships and state space constraints. Other types of control formulations, such as robust control, may also be considered specific forms of a stochastic program that are amenable to specific techniques to find control policies with given characteristics.

Continuous time stochastic models (see, e.g., Harrison [1985]) are also possible but generally require more simplified models than those considered in stochastic programming. Again, continuous time formulations are consistent with stochastic programs but have not been the main emphasis of research or the examples in this book. In certain examples again, they may be quite relevant (see, for example, Harrison and Wein [1990] for an excellent application in manufacturing) in defining fundamental solution characteristics, such as the optimality of control limit policies.

In all these control problems, the main emphasis is on characterizing solutions of some form of the dynamic programming Bellman-Hamilton-Jacobi equation or application of Pontryagin's maximum principle. Stochastic programs tend to view all decisions from beginning to end as part of the procedure. The dependence of the current decision on future outcomes and the transient nature of solutions are key elements. Section 3.5 provides some further explanation by describing these characteristics in terms of general optimality conditions.

e. Summary

Stochastic programming is simply another name for the study of optimal decision making under uncertainty. The term *stochastic programming* emphasizes a link to mathematical programming and algorithmic optimization procedures. These considerations dominate work in stochastic programming and distinguish stochastic programming from other fields of study. In this book, we follow this paradigm of concentrating on representation and characterizations of optimal decisions and on developing procedures to follow in determining optimal or approximately optimal decisions. This development begins in the next chapter with basic properties of stochastic program solution sets and optimal values.

Exercises

1. Consider the design problem in Section 1.4. Suppose the design decision does not completely specify x in (1.4.1), but the designer only knows that if a value $\hat{x}$ is specified then $x \in [.99\hat{x}, 1.01\hat{x}]$. Suppose a uniform distribution for x is assumed initially on this interval and that the designer can alter the design once after manufacturing and testing N axles out of a total predicted demand of 1,000 axles. The designer assumes that her posterior distribution on the actual mean relative to $\hat{x}$ would not change if she adjusts the target diameter $\hat{x}$ after observing the first N axle diameters. With these assumptions, formulate a Bayesian model to determine an initial specification $\hat{x}^1$ and N followed by a second specification $\hat{x}^2$ for the remaining $1000-N$ axles.
2. From the example in Section 1.2, suppose that a goal in each period is to realize a 16% return in each period with penalties $q=1$ and $r=4$ as before. Formulate the problem as in (10.2).
3. Consider the risk-sensitive model in (10.7) given initial state x^1, $\theta > 0$, $H=2$, and $\varepsilon^1 \sim N(\mu, \Sigma)$, the multivariate normal distribution with mean μ and variance-covariance matrix, Σ. Show that solving (10.7) is equivalent to solving the min-max problem:

$$\min_{u^1} \max_{\varepsilon^1} \theta [((u^1)^T R^1 u^1 + x^2(x^1, u^1, \varepsilon^1)^T Q^2 x^2(x^1, u^1, \varepsilon^1)^T) + (\varepsilon^1 - \mu)^T \Sigma^{-1} (\varepsilon^1 - \mu)], \quad (10.8)$$

   i.e., u^1 optimal in (10.8) is also optimal in (10.7) and vice versa as long as both problems have finite optimal values. To do this, first show that $\int e^{-Q(x,y)} dy = k e^{-\min_y Q(x,y)}$ for some constant k (independent of x) for any positive definite quadratic function $Q(x,y)$.

2.11 Short Reviews

a. Linear programming

Consider a linear program (LP) of the form

$$\max\{c^T x \mid Ax = b, x \geq 0\} , \tag{11.1}$$

where A is an $m \times n$ matrix, x and c are $n \times 1$ vectors, and b is an $m \times 1$ vector. If needed, any inequality constraint can be transformed into an equality by the addition of *slack variables*:

$$a_{i\cdot}x \leq b_i \quad \text{becomes} \quad a_{i\cdot}x + s_i = b_i ,$$

where s_i is the slack variable of row i and $a_{i\cdot}$ is the i th row of matrix A.

A *solution* to (11.1) is a vector x that satisfies $Ax = b$. A *feasible solution* is a solution x with $x \geq 0$. An *optimal solution* x^* is a feasible solution such that $c^T x^* \geq c^T x$ for all feasible solutions x. A *basis* is a choice of n linearly independent columns of A. Associated with a basis is a submatrix B of the corresponding columns, so that, after a suitable rearrangement, A can be partitioned into $A = [B, N]$. Associated with a basis is a *basic solution*, $x_B = B^{-1}b$, $x_N = 0$, and $z = c_B^T B^{-1} b$, where $[x_B, x_N]$ and $[c_B, c_N]$ are partitions of x and c following the basic and nonbasic columns. We use B^{-1} to denote the inverse of B, which is known to exist because B has linearly independent columns and is square.

In geometric terms, basic solutions correspond to *extreme points* of the polyhedron, $\{x \mid Ax = b, x \geq 0\}$. A basis is feasible (optimal) if its associated basic solution is feasible (optimal). The conditions for feasibility are $B^{-1}b \geq 0$. The conditions for optimality are that in addition to feasibility, the inequalities, $c_N^T - c_B^T B^{-1} N \leq 0$, hold.

Linear programs are routinely solved by widely distributed, easy-to-use LP solvers. Access to such a solver would be useful for some exercises in this book. For a better understanding, some examples and exercises also use manual solutions of linear programs.

Finding an optimal solution is equivalent to finding an optimal *dictionary*, a definition of individual variables in terms of the other variables. In the *simplex algorithm*, starting from a feasible dictionary, the next one is obtained by selecting an *entering variable* (any nonbasic variable whose increase leads to an increase in the objective value), then finding a *leaving variable* (the first to become negative as the entering variable increases), then realizing a *pivot* substituting the entering for the leaving variable in the dictionary. An optimal solution is reached when no entering variable can be found.

A linear program is *unbounded* if an entering variable exists for which no leaving variable can be found. In some cases, a feasible initial dictionary is not available at once. Then, *phase one* of the simplex method consists of finding such an initial dictionary. A number of artificial variables are introduced to make the dictionary feasible. The phase one procedure minimizes the sum of artificials using the simplex method. If a solution with a sum of artificials equal to zero exists, then the original problem is feasible and *phase two* continues with the true objective function. If the optimal solution of the phase one problem is nonzero, then the original problem is *infeasible*.

As an example, consider the following linear program:

$$\begin{aligned} &\max -x_1+3x_2 \\ &\text{s. t. } 2x_1+x_2 \geq 5 \ , \\ &\qquad x_1+x_2 \leq 3 \ , \\ &\qquad x_1,x_2 \geq 0 \ . \end{aligned}$$

Adding slack variables s_1 and s_2 , the two constraints read

$$\begin{aligned} 2x_1 + x_2 - s_1 &= 5 \ , \\ x_1 + x_2 + s_2 &= 3 \ . \end{aligned}$$

The natural choice for the initial basis is (s_1,s_2) . This basis is infeasible as s_1 would obtain the value -5 . An *artificial variable* (a_1) is added to row one to form:

$$2x_1+x_2-s_1+a_1=5 \ .$$

The phase-one problem consists of minimizing a_1 , i.e., finding $-\max -a_1$. Let $z=-a_1$ be the phase one objective, which after substituting for a_1 gives the initial dictionary in phase one:

$$\begin{aligned} z &= -5 + 2x_1 + x_2 - s_1 \ , \\ a_1 &= 5 - 2x_1 - x_2 + s_1 \ , \\ s_2 &= 3 - x_1 - x_2 \ , \end{aligned}$$

corresponding to the initial basis (a_1,s_2) . Entering candidates are x_1 and x_2 as they both increase the objective value. Choosing x_1 , the leaving variable is a_1 (because it becomes zero for $x_1=2.5$ while s_2 becomes zero only for $x_1=3$). Substituting x_1 for a_1 , the second dictionary becomes:

$$\begin{aligned} z &= -a_1 \ , \\ x_1 &= 2.5 - 0.5x_2 + 0.5s_1 - 0.5a_1 \ , \\ s_2 &= 0.5 - 0.5x_2 - 0.5s_1 + 0.5a_1 \ . \end{aligned}$$

This dictionary is an optimal dictionary for phase one. (No nonbasic variable would possibly increase x .) This means the original problem is feasible. (In fact, the basis (x_1,s_2) is feasible with solution $x_1=2.5$, $x_2=0.0$.)

We now turn to phase two. We replace the phase one objective with the original objective:

$$z=-x_1+3x_2=-2.5+3.5x_2-0.5s_1 \ .$$

By removing the artificial variable a_1 (as it is not needed anymore), we obtain the following first dictionary in phase two:

$$\begin{aligned} z &= -2.5 + 3.5x_2 - 0.5s_1 \ , \\ x_1 &= \ \ 2.5 \ - 0.5x_2 + 0.5s_1 \ , \\ s_2 &= \ \ 0.5 \ - 0.5x_2 - 0.5s_1 \ . \end{aligned}$$

The next entering variable is x_2 with leaving variable s_2. After substitution, we obtain the final dictionary:

$$\begin{aligned} z &= 1 - 4s_1 - 7s_2 \ , \\ x_1 &= 2 + \ s_1 \ + \ s_2 \ , \\ x_2 &= 1 - \ s_1 \ - 2s_2 \ , \end{aligned}$$

which is optimal because no nonbasic variable is a valid entering variable. The optimal solution is $x^* = (2,1)^T$ with $z^* = 1$.

b. Duality for linear programs

The *dual* of the so-called primal problem (11.1) is:

$$\min\{\pi^T b \mid \pi^T A \geq c^T, \pi \text{ unrestricted}\} \ . \tag{11.2}$$

Variables π are called *dual variables*. One such variable is associated with each constraint of the primal. When the primal constraint is an equality, the dual variable is *free* (unrestricted in sign). Dual variables are sometimes called *shadow prices* or *multipliers* (as in nonlinear programming). The dual variable π_i may sometimes be interpreted as the marginal value associated with resource b_i.

If the dual is unbounded, then the primal is infeasible. Similarly, if the primal is unbounded, then the dual is infeasible. Both problems can also be simultaneously infeasible.

If x is primal feasible and π is dual feasible, then $c^T x \leq \pi^T b$. The primal has an optimal solution x^* if and only if the dual has an optimal solution π^*. In that case, $c^T x^* = (\pi^*)^T b$ and the primal and dual solutions satisfy the *complementary slackness* conditions:

$$(a_{i\cdot})x^* = b_i \text{ or } \pi_i^* = 0 \text{ or both, for any } i = 1,\ldots,m \ ,$$

$$(\pi^*)^T a_{\cdot j} = c_j \text{ or } x_j^* = 0 \text{ or both, for any } j = 1,\ldots,n \ ,$$

where $a_{\cdot j}$ is the j-th column of A and, as before, $a_{i\cdot}$ is the i-th row of A.

An alternative presentation is to say that $s_i^* \pi_i^* = 0$, where s_i is the slack variable of the ith constraint, i.e., either the slack or the dual variable associated with a constraint is zero, and similarly for the second condition. Thus, the optimal solution of the dual can be recovered from the optimal solution for the primal, and vice versa.

The optimality conditions can also be interpreted to say that either there exists some *improving direction*, w , from a current feasible solution, $\hat{x}$, so that $c^T w > 0$, $w_j \geq 0$ for all $j \in N$, $N = \{j \mid \hat{x}_j = 0\}$, and $a_{i\cdot} w = 0$ for all $i \in I$, $I = \{i \mid a_{i\cdot}\hat{x} = b_i\}$ (hence, for $Ax = b$ in the primal system of (11.1), $I = \{1, \ldots, m\}$) or there exists some π such that $\sum_{i \in I} \pi_i a_{ij} \geq c_j$ for all $j \in N$, $\sum_{i \in I} \pi_i a_{ij} = c_j$ for all $j \notin N$, but both cannot occur. This result is equivalent to the *Farkas lemma*, which gives alternative systems with or without solutions.

The *dual simplex method* replicates on the primal solution what the iterations of the simplex method would be on the dual problem: it first finds the leaving variable (one that is strictly negative) then the entering variable (the first one that would become positive in the objective line). The dual simplex is particularly useful when a solution is already available to the original primal problem and some extra constraint or bound is added to the problem. The reader is referred to Chvátal [1980, pp. 152–157] for a detailed presentation.

Other material not covered in this section is meant to be restrictive to a given topic area. The next section discusses more of the mathematical properties of solutions and functions.

c. Nonlinear programming and convex analysis

When objectives and constraints may contain nonlinear functions, the optimization problem becomes a *nonlinear program*. The nonlinear program analogous to (11.1) has the form

$$\min\{f(x) \mid g(x) \leq 0, h(x) = 0\} , \tag{11.3}$$

where $x \in \Re^n$, $f : \Re^n \to \Re$, $g : \Re^n \to \Re^m$, and $h : \Re^n \to \Re^l$. We may also assume that the range of f may include ∞ to allow the objective to include constraints directly through an *indicator function*:

$$\delta(x \mid X) = \begin{cases} 0 & \text{if } g(x) \leq 0 \text{ , } h(x) = 0 \text{ ,} \\ +\infty & \text{otherwise,} \end{cases}$$

where X is the set of x satisfying the constraints in (11.3), i.e., the *feasible region*.

In this book, the feasible region is usually a *convex set* so that X contains any *convex combination*,

$$\sum_{i=1}^{s} \lambda^i x^i, \sum_{i=1}^{s} \lambda^i = 1, \lambda^i \geq 0 \text{ , } i = 1, \ldots, s \text{ ,}$$

of points, x^i , $i = 1, \ldots, s$, that are in the feasible region. Extreme points of the region are points that cannot be expressed as a convex combination of two distinct points also in the region. The set of all convex combinations of a given set of points is its *convex hull*.

The feasible region is also most generally *closed* so that it contains all limits of infinite sequences of points in the region. The region is also generally *connected*, so that, for any x^1 and x^2 in the region, there exists some path of points in the feasible region connecting x^1 to x^2 by a function, $\boldsymbol{\eta} : [0,1] \to \Re^n$ that is continuous with $\boldsymbol{\eta}(0) = x^1$ and $\boldsymbol{\eta}(1) = x^2$. For certain results, we may also assume the region is *bounded* so that a *ball* of radius M, $\{x \mid \|x\| \le M\}$, contains the entire set of feasible points. Otherwise, the region is *unbounded*. Note that a region may be unbounded while the optimal value in (11.1) or (11.3) is still bounded. In this case, the region often contains a *cone*, i.e., a set S such that if $x \in S$, then $\lambda x \in S$ for all $\lambda \ge 0$. When the region is both closed and bounded, then it is *compact*.

The set of equality constraints, $h(x) = 0$, is often *affine*, i.e., they can be expressed as linear combinations of the components of x and some constant. In this case, each constraint, $h_i(x) = 0$, is a *hyperplane*, $a_{i\cdot}x - b_i = 0$, as in the linear program constraints. In this case, $h(x) = 0$, defines an *affine space*, a *translation* of the *parallel subspace*, $Ax = 0$. The affine space *dimension* is the same as its parallel subspace, i.e., the maximum number of linearly independent vectors in the subspace.

With nonlinear constraints and inequalities, the region may not be an affine space, but we often consider the lowest-dimension affine space containing them, i.e., the *affine hull* of the region. The affine hull is useful in optimality conditions because it distinguishes *interior* points that can be the center of a ball entirely within the region from the *relative interior* (ri), which can be the center of a ball whose intersection with the affine hull is entirely within the region. When a point is not in a feasible region, we often take its *projection* into the region using an operator, Π. If the region is X, then the projection of x onto X is $\Pi(x) = \mathrm{argmin}\{\|w - x\| \mid w \in X\}$.

In this book, we generally assume that the objective function f is a *convex function*, i.e., such that

$$f(\lambda x^1 + (1-\lambda)x^2) \le \lambda f(x^1) + (1-\lambda)f(x^2),$$

$0 \le \lambda \le 1$. If f also is never $-\infty$ and is not $+\infty$ everywhere, then f is a *proper convex function*. The region where f is finite is called the *effective domain* of f ($\mathrm{dom}\ f$). We can also define convex functions in terms of the *epigraph* of f, $\mathrm{epi}\,(f) = \{(x,\beta) \mid \beta \ge f(x)\}$. In this case, f is convex if and only if its epigraph is convex. If $-f$ is convex, then f is *concave*.

Often, we assume that f has *directional derivatives*, $f'(x;w)$, that are defined as:

$$f'(x;w) = \lim_{\lambda \downarrow 0} \frac{f(x+\lambda w) - f(x)}{\lambda}\ .$$

When these limits exist and do not vary in all directions, then f is *differentiable*, i.e., there exists a *gradient*, ∇f, such that

$$\nabla f^T w = f'(x;w)$$

for all directions $w \in \Re^n$. We sometimes distinguish this standard form of differentiability from stricter forms as *Gâteaux* or *G-differentiability*. The stricter forms impose more conditions on the directional derivative such as uniform convergence over compact sets (*Hadamard* derivatives).

We also consider *Lipschitz* continuous or *Lipschitzian* functions such that $|f(x)-f(w)| \leq M\|x-w\|$ for any x and w and some $M < \infty$. If this property holds for all x and w in a set X, then f is Lipschitzian *relative to* X. When this property only holds *locally*, i.e., for $\|w-x\| \leq \varepsilon$ for some $\varepsilon > 0$, then f is *locally Lipschitz* at x.

Among differentiable functions, we often use *quadratic* functions that have a *Hessian* matrix of second derivatives, D, and can be written as

$$f(x) = c^T x + \frac{1}{2} x^T D x .$$

Many functions are not, however, differentiable. In this case, we express optimality in terms of *subgradients* at a point x, or vectors, η, such that

$$f(w) \geq f(x) + \eta^T (w-x)$$

for all w. In this case, $\{(x,\beta) \mid \beta = f(x) + \eta^T (w-x)\}$ is a *supporting hyperplane* of f at x. The set of subgradients at a point x is the *subdifferential* of f at x, written $\partial f(x)$.

Other useful properties include that f is *piecewise linear*, i.e., such that $f(x)$ is linear over regions defined by linear inequalities. When f is *separable* so that $f(x) = \sum_{i=1}^n f_i(x_i)$, then other advantages are possible in computation.

Given f convex and a convex feasible region in (11.3), we can define conditions that an optimal solution x^* and associated multipliers (π^*, ρ^*) must satisfy. In general, these conditions require some form of *regularity* condition. A common form is that there exists some $\hat{x}$ such that $g(\hat{x}) < 0$ and h is affine. This is generally called the *Slater condition*.

Given a regularity condition of this type, if the constraints in (11.3) define a feasible region, then x^* is optimal if and only if the *Karush-Kuhn-Tucker* conditions hold so that $x^* \in X$ and there exists $\pi^* \geq 0, \rho^*$ such that

$$\nabla f(x^*) + (\pi^*)^T \nabla g(x^*) + (\rho^*)^T \nabla h(x^*) = 0, \nabla g(x^*)^T \pi^* = 0 . \tag{11.4}$$

Optimality can also be expressed in terms of the *Lagrangian*:

$$l(x,\pi,\rho) = f(x) + \pi^T g(x) + \rho^T h(x) ,$$

so that sequentially minimizing over x and maximizing over π (in both orders) produces the result in (11.4). This occurs through a *Lagrangian dual problem* to (11.3) as

$$\max_{\pi \geq 0, \rho} \inf_x f(x) + \pi^T g(x) + \rho^T h(x) , \tag{11.5}$$

which is always a lower bound on the objective in (11.3) (*weak duality*), and, under the regularity conditions, yields equal optimal values in (11.3) and (11.4) (*strong duality*). In many cases, the Lagrangian can also be interpreted with the *conjugate* function of f, defined as

$$f^*(\pi) = \sup_x \{\pi^T x - f(x)\} ,$$

which is also a convex function if f is convex.

Our algorithms often apply to the Lagrangian to obtain *convergence*, i.e., a sequence of solutions, $x^\nu \to x^*$. In some cases, we also approximate the function so that $f^\nu \to f$ in some way. If this convergence is *pointwise*, then $f^\nu(x) \to f(x)$ for each x individually. If the convergence is *uniform* on a set X, then, for any $\varepsilon > 0$, there exists $N(\varepsilon)$ such that for all $\nu \geq N(\varepsilon)$ and all $x \in X$, $|f^\nu(x) - f(x)| < \varepsilon$.