## Contents

**Preface**  xiii

### 1 Stochastic Programming Models  1

1.1 Introduction .............................................. 1
1.2 Inventory .................................................. 1
   1.2.1 The Newsvendor Problem ............................. 1
   1.2.2 Chance Constraints .................................. 5
   1.2.3 Multistage Models .................................. 5
1.3 Multiproduct Assembly .................................. 8
   1.3.1 Two-Stage Model .................................. 8
   1.3.2 Chance Constrained Model ......................... 9
   1.3.3 Multistage Model ................................ 11
1.4 Portfolio Selection ................................... 12
   1.4.1 Static Model ..................................... 12
   1.4.2 Multistage Portfolio Selection .................. 15
   1.4.3 Decision Rules .................................. 18
Exercises .................................................... 20

### 2 Two-Stage Problems  21

2.1 Linear Two-Stage Problems ............................. 21
   2.1.1 Basic Properties .................................. 21
   2.1.2 The Expected Recourse Cost for Discrete Distributions 23
   2.1.3 The Expected Recourse Cost for General Distributions 26
   2.1.4 Optimality Conditions ............................ 31
2.2 Polyhedral Two-Stage Problems ...................... 35
   2.2.1 General Properties ................................ 35
   2.2.2 Expected Recourse Cost .......................... 36
   2.2.3 Optimality Conditions ............................ 39
2.3 General Two-Stage Problems .......................... 40
   2.3.1 Problem Formulation, Interchangeability .......... 40
   2.3.2 Convex Two-Stage Problems ...................... 41
2.4 Nonanticipativity ...................................... 44
   2.4.1 Scenario Formulation ................................ 44
   2.4.2 Dualization of Nonanticipativity Constraints .... 45
   2.4.3 Nonanticipativity Duality for General Distributions 47
   2.4.4 Value of Perfect Information .................... 50
Exercises .................................................... 51
## 3 Multistage Problems

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Problem Formulation</td>
<td>53</td>
</tr>
<tr>
<td>3.1.1 The General Setting</td>
<td>53</td>
</tr>
<tr>
<td>3.1.2 The Linear Case</td>
<td>58</td>
</tr>
<tr>
<td>3.1.3 Scenario Trees</td>
<td>61</td>
</tr>
<tr>
<td>3.1.4 Filtration Interpretation</td>
<td>63</td>
</tr>
<tr>
<td>3.1.5 Algebraic Formulation of Nonanticipativity Constraints</td>
<td>64</td>
</tr>
<tr>
<td>3.1.6 Piecewise Affine Policies</td>
<td>69</td>
</tr>
<tr>
<td>3.2 Duality</td>
<td>71</td>
</tr>
<tr>
<td>3.2.1 Convex Multistage Problems</td>
<td>71</td>
</tr>
<tr>
<td>3.2.2 Optimality Conditions</td>
<td>72</td>
</tr>
<tr>
<td>3.2.3 Dualization of Feasibility Constraints</td>
<td>75</td>
</tr>
<tr>
<td>3.2.4 Dualization of Nonanticipativity Constraints</td>
<td>76</td>
</tr>
<tr>
<td>Exercises</td>
<td>78</td>
</tr>
</tbody>
</table>

## 4 Optimization Models with Probabilistic Constraints

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>81</td>
</tr>
<tr>
<td>4.2 Convexity in Probabilistic Optimization</td>
<td>87</td>
</tr>
<tr>
<td>4.2.1 Generalized Concavity of Functions and Measures</td>
<td>87</td>
</tr>
<tr>
<td>4.2.2 Convexity of Probabilistically Constrained Sets</td>
<td>99</td>
</tr>
<tr>
<td>4.2.3 Connectedness of Probabilistically Constrained Sets</td>
<td>105</td>
</tr>
<tr>
<td>4.3 Separable Probabilistic Constraints</td>
<td>105</td>
</tr>
<tr>
<td>4.3.1 Continuity and Differentiability Properties of Distribution Functions</td>
<td>106</td>
</tr>
<tr>
<td>4.3.2 $p$-Efficient Points</td>
<td>108</td>
</tr>
<tr>
<td>4.3.3 The Tangent and Normal Cones of $\text{conv} \mathcal{Z}_p$</td>
<td>114</td>
</tr>
<tr>
<td>4.3.4 Optimality Conditions and Duality Theory</td>
<td>116</td>
</tr>
<tr>
<td>4.4 Optimization Problems with Nonseparable Probabilistic Constraints</td>
<td>130</td>
</tr>
<tr>
<td>4.4.1 Differentiability of Probability Functions and Optimality Conditions</td>
<td>131</td>
</tr>
<tr>
<td>4.4.2 Approximations of Nonseparable Probabilistic Constraints</td>
<td>134</td>
</tr>
<tr>
<td>4.5 Semi-Infinite Probabilistic Problems</td>
<td>141</td>
</tr>
<tr>
<td>Exercises</td>
<td>147</td>
</tr>
</tbody>
</table>

## 5 Statistical Inference

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Statistical Properties of SAA Estimators</td>
<td>151</td>
</tr>
<tr>
<td>5.1.1 Consistency of SAA Estimators</td>
<td>153</td>
</tr>
<tr>
<td>5.1.2 Asymptotics of the SAA Optimal Value</td>
<td>157</td>
</tr>
<tr>
<td>5.1.3 Second Order Asymptotics</td>
<td>160</td>
</tr>
<tr>
<td>5.1.4 Minimax Stochastic Programs</td>
<td>164</td>
</tr>
<tr>
<td>5.2 Stochastic Generalized Equations</td>
<td>168</td>
</tr>
<tr>
<td>5.2.1 Consistency of Solutions of the SAA Generalized Equations</td>
<td>169</td>
</tr>
<tr>
<td>5.2.2 Asymptotics of SAA Generalized Equations Estimators</td>
<td>171</td>
</tr>
<tr>
<td>5.3 Monte Carlo Sampling Methods</td>
<td>173</td>
</tr>
<tr>
<td>5.3.1 Exponential Rates of Convergence and Sample Size Estimates in the Case of a Finite Feasible Set</td>
<td>174</td>
</tr>
<tr>
<td>5.3.2 Sample Size Estimates in the General Case</td>
<td>178</td>
</tr>
<tr>
<td>5.3.3 Finite Exponential Convergence</td>
<td>183</td>
</tr>
<tr>
<td>5.4 Quasi–Monte Carlo Methods</td>
<td>185</td>
</tr>
<tr>
<td>5.5 Variance Reduction Techniques</td>
<td>190</td>
</tr>
</tbody>
</table>
## Contents

5.5.1 Latin Hypercube Sampling ............................................. 190
5.5.2 Linear Control Random Variables Method ......................... 191
5.5.3 Importance Sampling and Likelihood Ratio Methods ............. 192
5.6 Validation Analysis ......................................................... 193
5.6.1 Estimation of the Optimality Gap ................................. 194
5.6.2 Statistical Testing of Optimality Conditions ....................... 198
5.7 Chance Constrained Problems ........................................... 200
5.7.1 Monte Carlo Sampling Approach ................................... 201
5.7.2 Validation of an Optimal Solution ................................ 206
5.8 SAA Method Applied to Multistage Stochastic Programming ....... 210
5.8.1 Statistical Properties of Multistage SAA Estimators ............ 211
5.8.2 Complexity Estimates of Multistage Programs .................... 215
Exercises ........................................................................ 218

6 Risk Averse Optimization .................................................. 223
6.1 Introduction ................................................................. 223
6.2 Mean–Risk Models ......................................................... 224
6.2.1 Main ideas of mean–risk analysis .................................. 224
6.2.2 Semideviations .......................................................... 225
6.2.3 Weighted mean deviations from quantiles ....................... 226
6.2.4 Average Value-at-Risk ............................................... 227
6.3 Coherent Risk Measures .................................................. 231
6.3.1 Differentiability Properties of Risk Measures ................. 236
6.3.2 Examples of Risk Measures ......................................... 240
6.3.3 Law Invariant Risk Measures ...................................... 249
6.3.4 Spectral Risk Measures .............................................. 252
6.3.5 Kusuoka Representations ........................................... 256
6.3.6 Probability Spaces with Atoms .................................... 261
6.3.7 Stochastic Orders ...................................................... 263
6.4 Optimization of Risk Measures ......................................... 266
6.4.1 Dualization of Nonanticipativity Constraints ................. 268
6.4.2 Interchangeability Principle for Risk Measures ................ 270
6.4.3 Example ................................................................. 272
6.5 Multistage Risk Averse Optimization ................................ 276
6.5.1 Scenario Tree Formulation ......................................... 276
6.5.2 Conditional Risk Mappings ......................................... 282
6.5.3 Dynamic Risk Measures ............................................ 286
6.5.4 Risk Averse Multistage Stochastic Programming .............. 289
6.5.5 Time Consistency of Multi-period Problems .................... 292
6.5.6 Portfolio Selection and Inventory Model Examples ............ 298
Exercises ........................................................................ 301

7 Distributionally Robust Stochastic Programming .................... 307
7.1 Introduction ................................................................. 307
7.2 Construction of Ambiguity Sets of Probability Measures .......... 308
7.2.1 Distance Approach .................................................. 308
7.2.2 Approach of $\phi$-divergence ..................................... 313
7.3 Ambiguous Chance Constraints ....................................... 317
7.4 The Problem of Moments ............................................... 321
7.5 Statistical Properties of Distributionally Robust Functionals .... 326
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5.1 Average Value-at-Risk</td>
<td>326</td>
</tr>
<tr>
<td>7.5.2 Absolute Semideviation Risk Measure</td>
<td>331</td>
</tr>
<tr>
<td>7.5.3 Von Mises Statistical Functionals</td>
<td>333</td>
</tr>
<tr>
<td>7.5.4 Sample estimates of $\phi$-divergence functionals</td>
<td>336</td>
</tr>
<tr>
<td>7.6 Multistage Distributionally Robust Setting</td>
<td>338</td>
</tr>
<tr>
<td>7.6.1 Conditional Counterparts of Distributionally Robust Functionals</td>
<td>338</td>
</tr>
<tr>
<td>7.6.2 Composite Distributionally Robust Functionals</td>
<td>342</td>
</tr>
<tr>
<td>7.6.3 Optimal Control Model</td>
<td>348</td>
</tr>
<tr>
<td>7.6.4 Optimal Stopping Time</td>
<td>351</td>
</tr>
<tr>
<td>Exercises</td>
<td>358</td>
</tr>
</tbody>
</table>

### 8 Computational Methods

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 Introduction</td>
<td>361</td>
</tr>
<tr>
<td>8.2 Stochastic Approximation Method</td>
<td>362</td>
</tr>
<tr>
<td>8.2.1 Classical Approach</td>
<td>362</td>
</tr>
<tr>
<td>8.2.2 Robust SA Approach</td>
<td>365</td>
</tr>
<tr>
<td>8.2.3 Mirror Descent SA Method</td>
<td>367</td>
</tr>
<tr>
<td>8.2.4 Accuracy Certificates for Mirror Descent SA Solutions</td>
<td>374</td>
</tr>
<tr>
<td>8.3 Stochastic Dual Dynamic Programming Method</td>
<td>379</td>
</tr>
<tr>
<td>8.3.1 Approximate Dynamic Programming Approach</td>
<td>380</td>
</tr>
<tr>
<td>8.3.2 The SDDP algorithm</td>
<td>382</td>
</tr>
<tr>
<td>8.3.3 Convergence Properties of the SDDP Algorithm</td>
<td>385</td>
</tr>
<tr>
<td>8.3.4 Risk Averse SDDP Method</td>
<td>387</td>
</tr>
<tr>
<td>8.3.5 Duality Bounds</td>
<td>391</td>
</tr>
<tr>
<td>8.4 Periodical Multistage Programs</td>
<td>394</td>
</tr>
<tr>
<td>8.4.1 Wald–Bellman Equations</td>
<td>395</td>
</tr>
<tr>
<td>8.4.2 Sample Complexity</td>
<td>398</td>
</tr>
<tr>
<td>8.4.3 Cutting Planes Method</td>
<td>400</td>
</tr>
</tbody>
</table>

### 9 Background Material

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1 Optimization and Convex Analysis</td>
<td>403</td>
</tr>
<tr>
<td>9.1.1 Directional Differentiability</td>
<td>404</td>
</tr>
<tr>
<td>9.1.2 Elements of Convex Analysis</td>
<td>406</td>
</tr>
<tr>
<td>9.1.3 Optimization and Duality</td>
<td>410</td>
</tr>
<tr>
<td>9.1.4 Optimality Conditions</td>
<td>416</td>
</tr>
<tr>
<td>9.1.5 Perturbation Analysis</td>
<td>421</td>
</tr>
<tr>
<td>9.1.6 Epiconvergence</td>
<td>428</td>
</tr>
<tr>
<td>9.2 Probability</td>
<td>429</td>
</tr>
<tr>
<td>9.2.1 Probability Spaces and Random Variables</td>
<td>429</td>
</tr>
<tr>
<td>9.2.2 Conditional Probability and Conditional Expectation</td>
<td>434</td>
</tr>
<tr>
<td>9.2.3 Essential Supremum</td>
<td>436</td>
</tr>
<tr>
<td>9.2.4 Measurable Multifunctions and Random Functions</td>
<td>438</td>
</tr>
<tr>
<td>9.2.5 Expectation Functions</td>
<td>441</td>
</tr>
<tr>
<td>9.2.6 Uniform Laws of Large Numbers</td>
<td>446</td>
</tr>
<tr>
<td>9.2.7 Law of Large Numbers for Risk Measures</td>
<td>451</td>
</tr>
<tr>
<td>9.2.8 Law of Large Numbers for Random Sets and Subdifferentials</td>
<td>455</td>
</tr>
<tr>
<td>9.2.9 Delta Method</td>
<td>458</td>
</tr>
<tr>
<td>9.2.10 Exponential Bounds of the Large Deviations Theory</td>
<td>464</td>
</tr>
<tr>
<td>9.2.11 Uniform Exponential Bounds</td>
<td>470</td>
</tr>
<tr>
<td>9.3 Elements of Functional Analysis</td>
<td>475</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>9.3.1</td>
<td>Conjugate Duality and Differentiability</td>
</tr>
<tr>
<td>9.3.2</td>
<td>Paired Locally Convex Topological Vector Spaces</td>
</tr>
<tr>
<td>9.3.3</td>
<td>Lattice Structure</td>
</tr>
<tr>
<td>9.3.4</td>
<td>Interchangeability Principle</td>
</tr>
<tr>
<td>9.3.5</td>
<td>Hardy–Littlewood Inequality</td>
</tr>
<tr>
<td>Exercises</td>
<td></td>
</tr>
</tbody>
</table>

10 Bibliographical Remarks | 497  

Bibliography | 505  

Index | 521  |
Preface

Preface to the Third Edition

This is a substantial revision of the previous edition with added new material. The presentation of Chapter 6 is updated. In particular the Interchangeability Principle for risk measures is discussed in detail. Two new chapters are added. In Chapter 7 we present a systematic theory of distributionally robust stochastic programming (DRSP). Currently this is a hot topic of research. Particular attention is given to mathematical foundations of multistage formulations of DRSP. Statistical properties of empirical estimates of distributionally robust functionals are discussed. Time consistency of multistage problems is formulated in a general framework of preference systems with a particular application to distributionally robust stopping time problems. In Chapter 8 there is new material on formulation and numerical approaches to solving periodical multistage stochastic programs.

Preface to the Second Edition

In the second edition, we introduced new material to reflect recent developments in the area of stochastic programming. Chapter 6 underwent substantial revision. In sections 6.3.4–6.3.6, we extended the discussion of law invariant coherent risk measures and their Kusuoka representations. In sections 6.8.2–6.8.6, we provided in-depth analysis of dynamic risk measures and concepts of time consistency, including several new results.

In Chapter 4, we provided analytical description of the tangent and normal cones of chance constrained sets in section 4.3.3. We extended the analysis of optimality conditions in section 4.3.4 to nonconvex problems.

In Chapter 5, we added section 5.10 with a discussion of the stochastic dual dynamic programming method, which became popular in power generation planning. We also made corrections and small additions in Chapters 3 and 7, and we updated the bibliography.

Preface to the First Edition

The main topic of this book is optimization problems involving uncertain parameters, for which stochastic models are available. Although many ways have been proposed to model uncertain quantities, stochastic models have proved their flexibility and usefulness in diverse areas of science. This is mainly due to solid mathematical foundations and theoretical richness of the theory of probability and stochastic processes, and to sound statistical techniques of using real data.

Optimization problems involving stochastic models occur in almost all areas of science and engineering, as diverse as telecommunication, medicine, or finance, to name just a few. This stimulates interest in rigorous ways of formulating, analyzing, and solving such problems. Due
to the presence of random parameters in the model, the theory combines concepts of optimization
theory, the theory of probability and statistics, and functional analysis. Moreover, in recent years
the theory and methods of stochastic programming have undergone major advances. All these
factors motivated us to present in an accessible and rigorous form contemporary models and
ideas of stochastic programming. We hope that the book will encourage other researchers to
apply stochastic programming models and to undertake further studies of this fascinating and
rapidly developing area.

We do not try to provide a comprehensive presentation of all aspects of stochastic program-
ning, but we rather concentrate on theoretical foundations and recent advances in selected areas.
The book is organized in seven chapters. The first chapter addresses modeling issues. The basic
concepts, such as recourse actions, chance (probabilistic) constraints, and the nonanticipativity
principle, are introduced in the context of specific models. The discussion is aimed at providing
motivation for the theoretical developments in the book, rather than practical recommendations.

Chapters 2 and 3 present detailed development of the theory of two- and multistage stochastic
programming problems. We analyze properties of the models and develop optimality conditions
and duality theory in a rather general setting. Our analysis covers general distributions of uncer-
tain parameters, and also provides special results for discrete distributions, which are relevant for
numerical methods. Due to specific properties of two- and multistage stochastic programming
problems, we were able to derive many of these results without resorting to methods of functional
analysis.

The basic assumption in the modeling and technical developments is that the probability distri-
bution of the random data is not influenced by our actions (decisions). In some applications
this assumption could be unjustified. However, dependence of probability distribution on deci-
sions typically destroys the convex structure of the optimization problems considered, and our
analysis exploits convexity in a significant way.

Chapter 4 deals with chance (probabilistic) constraints, which appear naturally in many ap-
lications. The chapter presents the current state of the theory, focusing on the structure of the
problems, optimality theory, and duality. We present generalized convexity of functions and
measures, differentiability, and approximations of probability functions. Much attention is de-
voted to problems with separable chance constraints and problems with discrete distributions. We
also analyze problems with first order stochastic dominance constraints, which can be viewed as
problems with continuum of probabilistic constraints. Many of the presented results are relatively
new and were not previously available in standard textbooks.

Chapter 5 is devoted to statistical inference in stochastic programming. The starting point
of the analysis is that the probability distribution of the random data vector is approximated by
an empirical probability measure. Consequently the “true” (expected value) optimization prob-
lem is replaced by its sample average approximation (SAA). Origins of this statistical inference
go back to the classical theory of the maximum likelihood method routinely used in statistics.
Our motivation and applications are somewhat different, because we aim at solving stochastic
programming problems by Monte Carlo sampling techniques. That is, the sample is generated
in the computer and its size is only constrained by the computational resources needed to solve
the constructed SAA problem. One of the byproducts of this theory is the complexity analysis
of two- and multistage stochastic programming. Already in the case of two-stage stochastic pro-
gramming the number of scenarios (discretization points) grows exponentially with the increase
of the number of random parameters. Furthermore, for multistage problems, the computational
complexity also grows exponentially with the increase of the number of stages.

In Chapter 6 we outline the modern theory of risk averse approaches to stochastic program-
ning. We focus on the analysis of the models, optimality theory, and duality. Static and two-
stage risk averse models are analyzed in much detail. We also outline a risk averse approach to
multistage problems, using conditional risk mappings and the principle of “time consistency.”
Chapter 7 contains formulations of technical results used in the other parts of the book. For some of these less known results we give proofs, while for others we refer to the literature. The subject index can help the reader to find quickly a required definition or formulation of a needed technical result.

Several important aspects of stochastic programming have been left out. We do not discuss numerical methods for solving stochastic programming problems, with the exception of section 8.2 where the Stochastic Approximation method, and its relation to complexity estimates, is considered. Of course, numerical methods is an important topic which deserves careful analysis. This, however, is a vast and separate area which should be considered in a more general framework of modern optimization methods and to a large extent would lead outside the scope of this book.

We also decided not to include a thorough discussion of stochastic integer programming. The theory and methods of solving stochastic integer programming problems draw heavily from the theory of general integer programming. Their comprehensive presentation would entail discussion of many concepts and methods of this vast field, which would have little connection with the rest of the book.

At the beginning of each chapter we indicate the authors who were primarily responsible for writing the material, but the book is the creation of all three of us, and we share equal responsibility for errors and inaccuracies that escaped our attention.

We thank Stevens Institute of Technology and Rutgers University for granting sabbatical leaves to Darinka Dentcheva and Andrzej Ruszczyński, during which a large portion of this work was written. Andrzej Ruszczyński is also thankful to the Department of Operations Research and Financial Engineering of Princeton University for providing him with excellent conditions for his stay during the sabbatical leave.

*Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński*
Index

:= equal by definition, 403
$A^T$ transpose of matrix (vector) $A$, 403
$C(A')$ space of continuous functions, 159
$C^*$ polar of cone $C$, 406
$C^1(V, \mathbb{R}^n)$ space of continuously differentiable mappings, 170
$IF_\theta$ influence function, 334
$L_+^\perp$ orthogonal of (linear) space $L$, 34
$O(1)$ generic constant, 181
$O_p(\cdot)$ term, 458
$V_d(A)$ Lebesgue measure of set $A \subset \mathbb{R}^d$, 186
$W^{1,\infty}(U)$ space of Lipschitz continuous functions, 161, 424
$[a]_+ = \max\{a, 0\}$, 1
$I_A(\cdot)$ indicator function of set $A$, 404
$L_p(\Omega, \mathcal{F}, P)$ space, 478
$\Lambda(\bar{x})$ set of Lagrange multipliers vectors, 418
$N(\mu, \Sigma)$ normal distribution, 14
$N_C$ normal cone to set $C$, 407
$\Phi(z)$ cdf of standard normal distribution, 14
$\Pi_X$ metric projection onto set $X$, 362
$\Delta$ convergence in distribution, 158
$T^2(x, h)$ second order tangent set, 418
$AV@R$ Average Value-at-Risk, 228
$D(A, B)$ deviation of set $A$ from set $B$, 404
$D[Z_x]$ dispersion measure of random variable $Z_x$, 224
$E$ expectation, 430
$H(A, B)$ Hausdorff distance between sets $A$ and $B$, 404
$N$ set of positive integers, 429
$\mathbb{R}^n$ $n$-dimensional space, 403
$A$ domain of the conjugate of risk measure $\rho$, 232
$\mathcal{E}_n$ space of nonempty compact subsets of $\mathbb{R}^n$, 455
$\Delta$ set of probability density functions, 233, 236
$\mathcal{B}$ set of probability measures, 307, 308
$b(k; \alpha, N)$ cdf of binomial distribution, 204
$d$ distance generating function, 367
$g^+(x)$ right hand side derivative, 273
$\text{cl}(A)$ topological closure of set $A$, 404
$\text{conv}(A)$ convex hull of set $A$, 404
$\text{Corr}(X, Y)$ correlation of $X$ and $Y$, 191
$\text{Cov}(X, Y)$ covariance of $X$ and $Y$, 174
$q_\alpha$ weighted mean deviation, 226
$\mathcal{S}_C(\cdot)$ support function of set $C$, 407
$\delta(\omega)$ measure of mass one at the point $\omega$, 432
$\text{dist}(x, A)$ distance from point $x$ to set $A$, 404
$\text{dom} f$ domain of function $f$, 403
$\text{dom} \mathcal{G}$ domain of multifunction $\mathcal{G}$, 438
$\mathbb{R}$ set of extended real numbers, 403
$\text{epi} f$ epigraph of function $f$, 403
$\Rightarrow$ epiconvergence, 449
$\text{Exp}(K)$ set of exposed point of the set $K$, 477
$\text{Ext}(\Omega)$ set of extreme points of the set $\Omega$, 325
$\hat{S}_N$ the set of optimal solutions of the SAA problem, 152
$\hat{S}_N^\varepsilon$ the set of $\varepsilon$-optimal solutions of the SAA problem, 174
$\hat{\theta}_N$ optimal value of the SAA problem, 152
$\hat{f}_N(x)$ sample average function, 151
$L_A(\cdot)$ characteristic function of set $A$, 404
$\text{int}(C)$ interior of set $C$, 406
$|a|$ integer part of $a \in \mathbb{R}$, 209
$lsc f$ lower semicontinuous hull of function $f$, 403
$\mathcal{R}_C$ radial cone to set $C$, 408
$T_C$ tangent cone to set $C$, 408
$\nabla^2 f(x)$ Hessian matrix of second order partial derivatives, 172
$\partial$ subdifferential, 408
$\partial^\circ$ Clarke generalized gradient, 406
$\partial \varepsilon$ epsilon subdifferential, 456
$\text{pos} W$ positive hull of matrix $W$, 23
$\preceq_C$ partial order defined by cone $C$, 42
$\text{Pr}(A)$ probability of event $A$, 430
$\text{ri}$ relative interior, 406
$S^\varepsilon$ the set of $\varepsilon$-optimal solutions of the true problem, 174
$\sigma(\xi_1, \ldots, \xi_t)$ subalgebra generated by $(\xi_1, \ldots, \xi_t)$, 67
$\sigma^+_p$ upper semideviation, 225
$\sigma^-_p$ lower semideviation, 225
set of spectral functions, 252
\(\text{Tr}(A)\) trace of a square matrix, 411
\(\text{V@R}_\alpha\) Value-at-Risk, 226
\(\text{Var}[X]\) variance of \(X\), 13
\(|\Omega|\) cardinality of (finite) set \(\Omega\), 62
\(\vartheta^*\) optimal value of the true problem, 152
\(\xi_{[t]} = (\xi_1, \ldots, \xi_t)\) history of the process, 53
\(a \vee b = \max\{a, b\}\), 179
\(f^*(x, d)\) generalized directional derivative, 406
\(g'(x, h)\) directional derivative, 404
\(o_p(\cdot)\) term, 458
\(p\)-efficient point, 108
\(\text{iid}\) independently identically distributed, 152
ambiguity set, 307
approximation
  conservative, 227
Average Value-at-Risk, 227–229, 242
dual representation, 483
Banach lattice, 484
Banach space, 476
reflexive, 476
separable, 476
bilinear form, 232, 476
Borel set, 429
bounded in probability, 458
bracket, 463
bracketing number, 463
Bregman divergence, 368
capacity expansion, 25, 34, 50
Cauchy sequence, 475
chain rule, 459
chance constrained problem
  ambiguous, 317
  disjunctive semi-infinite formulation, 109
chance constraints, 5, 10, 14, 200
Clarke generalized gradient, 406
CLT Central Limit Theorem, 140
common random number generation method, 173
comonotonic random variables, 259
complexity
  of multistage programs, 215
  of two-stage programs, 174, 180
conditional expectation, 55, 434
conditional probability, 434
conditional risk mapping, 278, 283
conditional sampling, 210
  identical, 211
  independent, 210
Conditional Value-at-Risk, 227, 229
cone
  contingent, 416, 461
  critical, 171, 418
  normal, 407
  pointed, 483
  polar, 23
  recession, 23
  tangent, 408
confidence interval, 158
conjugate duality, 411, 481
constraint
  nonanticipativity, 44, 268
constraint qualification
  linear independence, 163, 172
  Mangasarian–Fromovitz, 417
  Robinson, 417
  Slater, 157
contact point, 477, 479
  contingent cone, 416, 461
  convergence
    in distribution, 158, 458
    in probability, 458
    weak, 460
  with probability one, 446
convex hull, 404
cumulative distribution function, 2
  of random vector, 10
cutting plane, 379
decision rule, 19, 54
Delta Theorem, 460, 461
  finite dimensional, 459
  second-order, 462
deterministic equivalent, 24, 41, 361
deviation of a set, 404
diameter
  of a set, 179
differential uniform dominance condition, 142
directional derivative, 404
\(\varepsilon\)-directional derivative, 457
generalized, 406
Hadamard, 459
second order, 461
tangentially to a set, 462
distance
  probability measures, 309
  total variation, 311
distribution
  asymptotically normal, 158
  binomial, 466
  conditional, 434
  Dirichlet, 91
  discrete, 431
  discrete with a finite support, 431
  empirical, 151
  Gamma, 94
  log-concave, 90
  log-normal, 99
  multivariate normal, 14, 89
  multivariate Student, 147
  normal, 158
  Pareto, 148
  uniform, 89
  Wishart, 95
distributionally robust functional, 307
  composite, 342
  conditional, 339
domain
  of multifunction, 438
  of a function, 403
  Donsker class, 463
dual feasibility condition, 124
duality gap, 410, 411
dynamic programming equations, 6, 54, 280
deterministic equivalent, 24, 41, 361
deviation of a set, 404
diameter
  of a set, 179
differential uniform dominance condition, 142
directional derivative, 404
\(\varepsilon\)-directional derivative, 457
generalized, 406
Hadamard, 459
second order, 461
tangentially to a set, 462
distance
  probability measures, 309
  total variation, 311
distribution
  asymptotically normal, 158
  binomial, 466
  conditional, 434
  Dirichlet, 91
  discrete, 431
  discrete with a finite support, 431
  empirical, 151
  Gamma, 94
  log-concave, 90
  log-normal, 99
  multivariate normal, 14, 89
  multivariate Student, 147
  normal, 158
  Pareto, 148
  uniform, 89
  Wishart, 95
distributionally robust functional, 307
  composite, 342
  conditional, 339
domain
  of multifunction, 438
  of a function, 403
  Donsker class, 463
dual feasibility condition, 124
duality gap, 410, 411
dynamic programming equations, 6, 54, 280
empirical cdf, 3
empirical distribution, 151
total variation, 311
distribution
  asymptotically normal, 158
  binomial, 466
  conditional, 434
  Dirichlet, 91
  discrete, 431
  discrete with a finite support, 431
  empirical, 151
  Gamma, 94
  log-concave, 90
  log-normal, 99
  multivariate normal, 14, 89
  multivariate Student, 147
  normal, 158
  Pareto, 148
  uniform, 89
  Wishart, 95
distributionally robust functional, 307
  composite, 342
  conditional, 339
domain
  of multifunction, 438
  of a function, 403
  Donsker class, 463
dual feasibility condition, 124
duality gap, 410, 411
dynamic programming equations, 6, 54, 280
empirical cdf, 3
empirical distribution, 151
entropy function, 368
epiconvergence, 428
  with probability one, 449
epigraph of a function, 403
epislon subdifferential, 456
essential supremum, 437
estimator
  common random number, 196
  consistent, 153
<table>
<thead>
<tr>
<th>term</th>
<th>page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear control</td>
<td>191</td>
</tr>
<tr>
<td>unbiased</td>
<td>152</td>
</tr>
<tr>
<td>expected value</td>
<td>430</td>
</tr>
<tr>
<td>well defined</td>
<td>431</td>
</tr>
<tr>
<td>expected value of perfect information</td>
<td>51</td>
</tr>
<tr>
<td>Fatou’s lemma</td>
<td>431</td>
</tr>
<tr>
<td>filtration</td>
<td>63, 67, 276, 289</td>
</tr>
<tr>
<td>floating body of a probability measure</td>
<td>97</td>
</tr>
<tr>
<td>Fréchet differentiability</td>
<td>404</td>
</tr>
<tr>
<td>(\alpha)-concave</td>
<td>87</td>
</tr>
<tr>
<td>(\alpha)-concave on a set</td>
<td>97</td>
</tr>
<tr>
<td>affine</td>
<td>70</td>
</tr>
<tr>
<td>biconjugate</td>
<td>232, 479</td>
</tr>
<tr>
<td>Carathéodory</td>
<td>152, 164, 439</td>
</tr>
<tr>
<td>characteristic</td>
<td>404</td>
</tr>
<tr>
<td>Clarke-regular</td>
<td>95, 406</td>
</tr>
<tr>
<td>composite</td>
<td>237</td>
</tr>
<tr>
<td>conjugate</td>
<td>232, 408, 479</td>
</tr>
<tr>
<td>continuously differentiable</td>
<td>406</td>
</tr>
<tr>
<td>cost-to-go</td>
<td>55, 59, 281</td>
</tr>
<tr>
<td>cumulative distribution (cdf)</td>
<td>2, 430</td>
</tr>
<tr>
<td>distance generating</td>
<td>367</td>
</tr>
<tr>
<td>disutility</td>
<td>223, 241</td>
</tr>
<tr>
<td>essentially bounded</td>
<td>478</td>
</tr>
<tr>
<td>extended real valued</td>
<td>430</td>
</tr>
<tr>
<td>indicator</td>
<td>22, 404</td>
</tr>
<tr>
<td>influence</td>
<td>334</td>
</tr>
<tr>
<td>integrable</td>
<td>431</td>
</tr>
<tr>
<td>likelihood ratio</td>
<td>192</td>
</tr>
<tr>
<td>log-concave</td>
<td>87</td>
</tr>
<tr>
<td>logarithmically concave</td>
<td>87</td>
</tr>
<tr>
<td>lower semicontinuous</td>
<td>403</td>
</tr>
<tr>
<td>moment generating</td>
<td>464</td>
</tr>
<tr>
<td>monotone</td>
<td>484</td>
</tr>
<tr>
<td>optimal value</td>
<td>439</td>
</tr>
<tr>
<td>polyhedral</td>
<td>22, 35, 403, 494</td>
</tr>
<tr>
<td>proper</td>
<td>403</td>
</tr>
<tr>
<td>quasi-concave</td>
<td>89</td>
</tr>
<tr>
<td>radical-inverse</td>
<td>188</td>
</tr>
<tr>
<td>random</td>
<td>439</td>
</tr>
<tr>
<td>random lower</td>
<td></td>
</tr>
<tr>
<td>semicontinuous</td>
<td>439</td>
</tr>
<tr>
<td>random polyhedral</td>
<td>35</td>
</tr>
<tr>
<td>sample average</td>
<td>447</td>
</tr>
<tr>
<td>spectral</td>
<td>252</td>
</tr>
<tr>
<td>strongly convex</td>
<td>409</td>
</tr>
<tr>
<td>subdifferentiable</td>
<td>408, 480</td>
</tr>
<tr>
<td>sublinear</td>
<td>476</td>
</tr>
<tr>
<td>support</td>
<td>407</td>
</tr>
<tr>
<td>utility</td>
<td>224, 241</td>
</tr>
<tr>
<td>well defined</td>
<td>441</td>
</tr>
<tr>
<td>Gâteaux differentiability</td>
<td>404, 459</td>
</tr>
<tr>
<td>generalized equation</td>
<td></td>
</tr>
<tr>
<td>sample average approximation</td>
<td>169</td>
</tr>
<tr>
<td>generic value</td>
<td>431</td>
</tr>
<tr>
<td>Hadamard differentiability</td>
<td>459</td>
</tr>
<tr>
<td>Hausdorff distance</td>
<td>404</td>
</tr>
<tr>
<td>here-and-now solution</td>
<td>9</td>
</tr>
<tr>
<td>Hessian matrix</td>
<td>418</td>
</tr>
<tr>
<td>higher order distribution functions</td>
<td>84</td>
</tr>
<tr>
<td>Hoffman’s lemma</td>
<td>414</td>
</tr>
<tr>
<td>identically distributed</td>
<td>446</td>
</tr>
<tr>
<td>importance sampling</td>
<td>193</td>
</tr>
<tr>
<td>independent identically distributed</td>
<td>446</td>
</tr>
<tr>
<td>inequality</td>
<td></td>
</tr>
<tr>
<td>Chebyshev</td>
<td>433</td>
</tr>
<tr>
<td>Chernoff</td>
<td>467</td>
</tr>
<tr>
<td>Hölder</td>
<td>479</td>
</tr>
<tr>
<td>Hardy–Littlewood</td>
<td>490</td>
</tr>
<tr>
<td>Hoeffding</td>
<td>466</td>
</tr>
<tr>
<td>Jensen</td>
<td>433</td>
</tr>
<tr>
<td>Markov</td>
<td>433</td>
</tr>
<tr>
<td>Minkowski for matrices</td>
<td>93</td>
</tr>
<tr>
<td>inf-compactness condition</td>
<td>154, 421</td>
</tr>
<tr>
<td>interchangeability principle</td>
<td>485, 486</td>
</tr>
<tr>
<td>for risk measures</td>
<td>270</td>
</tr>
<tr>
<td>for two-stage programming</td>
<td>41</td>
</tr>
<tr>
<td>interior of a set</td>
<td>406</td>
</tr>
<tr>
<td>inventory model</td>
<td>1, 272</td>
</tr>
<tr>
<td>Jacobian matrix</td>
<td>405</td>
</tr>
<tr>
<td>Kelley’s cutting plane</td>
<td></td>
</tr>
<tr>
<td>algorithm</td>
<td>385</td>
</tr>
<tr>
<td>Kullback–Leibler divergence</td>
<td>315, 337</td>
</tr>
<tr>
<td>Kusuoka representation</td>
<td>258</td>
</tr>
<tr>
<td>Lagrange multiplier</td>
<td>418</td>
</tr>
<tr>
<td>large deviations rate function</td>
<td>464</td>
</tr>
<tr>
<td>lattice</td>
<td>484</td>
</tr>
<tr>
<td>law invariance</td>
<td>308</td>
</tr>
<tr>
<td>Law of Large Numbers</td>
<td>2, 446</td>
</tr>
<tr>
<td>for random sets</td>
<td>455</td>
</tr>
<tr>
<td>pointwise</td>
<td>447</td>
</tr>
<tr>
<td>strong</td>
<td>446</td>
</tr>
<tr>
<td>uniform</td>
<td>447</td>
</tr>
<tr>
<td>weak</td>
<td>446</td>
</tr>
<tr>
<td>least upper bound</td>
<td>483</td>
</tr>
<tr>
<td>Lebesgue–Stieltjes integral</td>
<td>256</td>
</tr>
<tr>
<td>Lindeberg condition</td>
<td>141</td>
</tr>
<tr>
<td>Lipschitz continuous</td>
<td>405</td>
</tr>
<tr>
<td>lower bound</td>
<td></td>
</tr>
<tr>
<td>statistical</td>
<td></td>
</tr>
<tr>
<td>Lyapunov condition</td>
<td>140</td>
</tr>
<tr>
<td>mapping</td>
<td></td>
</tr>
<tr>
<td>contracting</td>
<td>475</td>
</tr>
<tr>
<td>convex</td>
<td>42</td>
</tr>
<tr>
<td>measurable</td>
<td>430</td>
</tr>
<tr>
<td>Markov chain</td>
<td>62</td>
</tr>
<tr>
<td>Markovian process</td>
<td>53</td>
</tr>
<tr>
<td>mean absolute deviation</td>
<td>225</td>
</tr>
<tr>
<td>measurable multifunction</td>
<td>438</td>
</tr>
<tr>
<td>measurable selection</td>
<td>439</td>
</tr>
<tr>
<td>measure</td>
<td></td>
</tr>
<tr>
<td>(\alpha)-concave</td>
<td>89</td>
</tr>
<tr>
<td>(\sigma)-finite</td>
<td>429</td>
</tr>
<tr>
<td>absolutely continuous</td>
<td>429</td>
</tr>
<tr>
<td>complete</td>
<td>429</td>
</tr>
<tr>
<td>Dirac</td>
<td>432</td>
</tr>
<tr>
<td>finite</td>
<td>429</td>
</tr>
<tr>
<td>Lebesgue</td>
<td>429</td>
</tr>
<tr>
<td>nonatomic</td>
<td>440</td>
</tr>
<tr>
<td>product</td>
<td>433</td>
</tr>
<tr>
<td>sigma-additive</td>
<td>429</td>
</tr>
<tr>
<td>metric projection</td>
<td>362</td>
</tr>
<tr>
<td>metric space</td>
<td>437</td>
</tr>
<tr>
<td>complete</td>
<td>475</td>
</tr>
<tr>
<td>Mirror Descent SA</td>
<td>371</td>
</tr>
<tr>
<td>model state equations</td>
<td>60</td>
</tr>
<tr>
<td>model state variables</td>
<td>60</td>
</tr>
<tr>
<td>moment generating function</td>
<td>464</td>
</tr>
<tr>
<td>multifunction</td>
<td>438</td>
</tr>
<tr>
<td>closed, 169, 421, 438</td>
<td></td>
</tr>
<tr>
<td>closed valued</td>
<td>169, 438</td>
</tr>
<tr>
<td>convex</td>
<td>41</td>
</tr>
<tr>
<td>convex valued</td>
<td>41, 440</td>
</tr>
<tr>
<td>measurable</td>
<td>438</td>
</tr>
<tr>
<td>optimal solution</td>
<td>439</td>
</tr>
<tr>
<td>upper semicontinuous</td>
<td>456</td>
</tr>
<tr>
<td>newsvendor problem</td>
<td>1, 303</td>
</tr>
<tr>
<td>node</td>
<td></td>
</tr>
<tr>
<td>ancestor</td>
<td>61</td>
</tr>
</tbody>
</table>
children, 61
root, 61
nonanticipativity, 6, 44, 53
nonanticipativity constraints, 65, 279
nonatomic probability space, 440

norm
  dual, 367, 476
normal cone, 407
normal integrands, 439

optimal stopping time
  optimistic, 356
  risk averse, 356
  risk neutral, 351

optimality conditions
  first order, 198, 416
  Karush–Kuhn–Tucker (KKT), 168, 198, 418
  second order, 172, 418
orbit, 250, 492
orthogonal measures, 432

paired spaces, 482
partial order, 436, 483
partition of a set, 435
point
  contact, 477, 479
  exposed, 477
  extreme, 325
  saddle, 410
polar cone, 406

policy
  basestock, 7, 301
  feasible, 7, 15, 54
  fixed mix, 18
  implementable, 7, 15, 54
  myopic, 17, 299
  optimal, 7, 55, 59
  piecewise affine, 70, 71
  piecewise linear, 70
  time consistent, 296, 351
portfolio selection, 12, 275
positive hull, 23
positively homogeneous, 171
Preface, xiii

preference system, 352
dynamically consistent, 352
recursive, 352
principle of conditional optimality, 293

probabilistic constraints, 5, 10, 81, 157
  individual, 83

joint, 83
probabilistic liquidity constraint, 87
probability density function, 430
probability distribution, 430
probability measure, 429
probability space, 429
  nonatomic, 440
  standard, 249
probability vector, 277

problem
  chance constrained, 81, 200
distributionally robust, 321
  first stage, 9
  of moments, 321
  piecewise linear, 184
  risk averse, 387
  risk neutral, 387
  second stage, 9

semi-infinite programming, 324
subconsistent, 411
time consistent, 296
two-stage, 9
weakly time consistent, 296

process
  autoregressive, 57
  stagewise independent, 53
  stochastic, 53
  product measure, 433
  prox-function, 368
  prox-mapping, 368

quadratic growth condition, 183, 420
quantile, 14
  left side, 3, 226, 250
  right side, 3, 226, 250
radial cone, 408
random function
  convex, 442
random variable, 430
  subexponential, 469
  subgaussian, 468
random vector, 430
recession cone, 407
recourse
  complete, 26
  fixed, 26, 37
  relatively complete, 9, 27, 55
  simple, 26
recourse action, 2
rectangularity, 344
relative interior, 406
risk mapping
  coherent conditional, 278, 283
cvx conditional, 278, 283
risk measure, 231
  absolute semideviation, 331, 359
axioms, 231
  coherent, 231
comonotonic, 259
consistent with stochastic orders, 264
cvx, 231
distribution based, 249
dynamic, 286, 351
dynamically consistent, 287
entropic, 244, 316
essential supremum, 229, 236, 259
law invariant, 249
mean-deviation, 246
mean-upper-semideviation, 247
mean-upper-semideviation from a target, 248
mean-variance, 245
nested, 280, 290
proper, 231
regular, 262
spectral, 254
strictly dynamically consistent, 295
version independent, 249
robust optimization, 10

saddle point, 410
sample
  independently identically distributed (iid), 152
random, 151
sample average approximation (SAA), 151
multistage, 210
sample covariance matrix, 199
sampling
  Latin Hypercube, 190
  Monte Carlo, 173
scenario tree, 61
scenarios, 3, 24
SDDP algorithm, 382
second order regularity, 419
second order tangent set, 418
semi-infinite probabilistic problem, 141
semideviation
  lower, 225
  upper, 225
separable space, 460
sequence
  Halton, 189
  log-concave, 98
  low-discrepancy, 188
  van der Corput, 189
set
  dual, 233
  elementary, 429
  generating, 253
  law invariant, 251
  polar, 483
  symmetric, 309, 483
Shapley–Folkman lemma, 407
sigma algebra, 429
  Borel, 429
  generated by countable
    partition, 436
  trivial, 429
significance level, 5
simplex, 368
Slater condition, 157
Snell envelope, 355, 357
solution
  ε-optimal, 174
  sharp, 182, 184
space
  Banach, 476
  decomposable, 485
dual, 476
  Hilbert, 245
  measurable, 429
  paired, 482
  probability, 429
  reflexive, 476
  sample, 429
stagewise independence, 7, 53, 62
standard probability space, 249
star discrepancy, 187
stationary point of α-concave
  function, 96
Stochastic Approximation, 362
stochastic dominance
  kth order, 84
  first order, 84, 264
  higher order, 84
  second order, 264
stochastic dominance
  constraint, 85
stochastic generalized
  equations, 168
stochastic order, 84, 263
  increasing convex, 253, 264
  usual, 264
stochastic programming
  nested risk averse multistage, 279, 290
stochastic programming problem
  minimax, 164
  multiperiod, 58
  multistage, 54
  multistage linear, 60
  two-stage convex, 41
two-stage linear, 21
two-stage polyhedral, 35
stochastic-order constraint, 85
stopping time, 351
strict complementarity
  condition, 173, 200
strict dynamic consistency, 352
strict monotonicity, 255, 488
  conditional, 295, 340
strongly regular solution of a
  generalized equation, 170
subdifferential, 408, 480
subgradient, 408, 480
  algebraic, 480
  stochastic, 362
support
  of a measure, 29, 430
  of a set, 407
support function, 22, 407, 408
  supporting plane, 379
tangent cone, 408
theorem
  Artstein–Vitale, 455
  Artstein–Wets, 449
  Aumann, 441
  Banach fixed point, 475
  Banach–Alaoglu, 477
  Birkhoff, 103
  Borell, 90
  Central Limit, 140
  Cramér’s Large Deviations, 464
  Danskin, 422
  Donsker, 463
  Fenchel–Moreau, 232, 409, 479
  Fubini, 433
  functional CLT, 159
  Glivenko–Cantelli, 448, 451
  Hahn–Banach, 476
  Helly, 407
  Hlawka, 188
  Klee–Nachbin–Namioka, 484
  Koksma, 187
  Kusuoka, 257
  Lebesgue decomposition, 432
  Lebesgue dominated
    convergence, 431
  Levin–Valadier, 423
  Lyapunov, 441
  Mazur, 477
  measurable selection, 439
  Minkowski, 325
  monotone convergence, 431
  Moreau–Rockafellar, 408, 481
  Rademacher, 406, 424
  Radon–Nikodym, 429
  Richter–Rogosinski, 432
  Riesz representation, 482
  Skorohod–Dudley almost
    sure representation, 460
time consistency
  of problem formulation, 293
  policy, 296
topology
  strong (norm), 476
  weak, 476
  weak∗, 476
total variation norm, 311, 482
uncertainty set, 10
uniformly integrable, 431, 458
upper bound
  conservative, 195
  statistical, 195
utility model, 241
value function, 6
Value-at-Risk, 15, 226, 243
  constraint, 15
variation
  of a function, 187
  variational inequality, 168
  stochastic, 168
von Mises statistical functional, 333
wait-and-see solution, 9, 50
Wald–Bellman (WB) equations, 396
Wasserstein distance, 312
weighted mean deviation, 226