Preface

Scope

In an abstract form, statistical decision making is an optimization problem that uses available statistical data as an input and optimizes an objective function of interest with respect to decision variables subject to certain constraints. Typically, uncertainty encoded in statistical data can be translated into five basic notions: likelihood, entropy, error, deviation, and risk. As a result, the majority of statistical decision problems can be tentatively divided into four major categories: (i) likelihood maximization, (ii) entropy maximization (relative entropy minimization), (iii) error minimization (regression), and (iv) decision models in which deviation or risk is either minimized or constrained. All these problems may include so-called technical constrains on decision variables, e.g., box and cardinality constraints. It is also common to optimize one of the corresponding five functionals while to constrain another, e.g., maximizing entropy subject to a constraint on deviation, or to find a trade-off between one of the functionals and the expected value of the quantity of interest. The book aims to demonstrate how to use these "building blocks": likelihood, entropy, error, deviation, and risk to formulate statistical decision problems arising in various risk management applications, e.g., optimal hedging, portfolio optimization, portfolio replication, cash flow matching, and classification and how to solve those problems in optimization package Portfolio Safeguard (PSG).

Content

The book consists of three parts: selected concepts of statistical decision theory (Part I), statistical decision problems (Part II), and case studies with PSG (Part III). Part I presents a general theory of error, deviation, and risk measures to be used in various statistical decision problems and also discusses probabilistic inequalities with deviation measures such as generalized Chebyshev's and Kolmogorov's

vi Preface

inequalities. Part II covers five major topics: parametric and nonparametric estimation based on the maximum likelihood principle, entropy maximization problems, unconstrained and constrained linear regression with general error measures, classification with logistic regression and with support vector machines (SVMs), and statistical decision problems with general deviation and risk measures. Part III discusses 21 case studies of typical statistical decision problems arising in risk management, particularly in financial engineering, and demonstrates implementation of those problems in PSG. All case studies are closely related to theoretical Part II and are examples of statistical decision problems from the four categories (i)–(iv).

Audience

The book is aimed at practitioners working in the areas of risk management, decision making under uncertainty, and statistics. It can serve as a quick introduction into the theory of general error, deviation, and risk measures for the graduate students, engineers, and statisticians interested in modeling and managing risk in various applications such as optimal hedging, portfolio replication, portfolio optimization, cash flow matching, structuring of collateralized debt obligations (CDOs), classification, sparse signal reconstruction, and therapy treatment optimization, to mention just a few. It can also be used as a supplementary reading for a number of graduate courses including but not limited to those of statistical analysis, models of risk, data mining, stochastic programming, financial engineering, modern portfolio theory, and advanced engineering economy.

Optimization Software: Portfolio Safeguard

PSG is an advanced nonlinear mixed-integer optimization package for solving a wide range of optimization, statistics, and risk management problems. PSG is a product of American Optimal Decisions, Inc. (see www.aorda.com). Although PSG is a general-purpose decision support tool, the focus application areas are risk management, financial engineering, military, and medical applications. PSG is based on a simple but powerful idea: for every engineering area, identify most commonly used nonlinear functions and include them in the package as independent built-in objects. Each function is defined by a function type, parameters, and a matrix of data (e.g., scenario matrix or covariance matrix). Specialized algorithms, built for different types of functions, efficiently optimize large-scale nonlinear functions, such as probability, value-at-risk (VaR), and omega functions, which are typically beyond the scope of commercial packages. The built-in function library provides simple and convenient interface for evaluating functions and their derivatives, for constructing optimization problems and solving them, and for analyzing solutions. No programming experience is required to use PSG.

PSG operates in four programming environments: Shell (Windows), MATLAB, C++, and Run-File (Text). The standard PSG setup includes case studies from

Preface vii

various areas with emphasis on financial engineering applications, such as portfolio optimization, asset allocation, selection of insurance, hedging with derivative contracts, bond matching, and structuring of CDOs.

PSG can be downloaded from the American Optimal Decisions web site: www. aorda.com/aod. Four types of licenses are available: Freeware Express, Regular, Academic, and Regular Business. Freeware Express Edition limits the number of decision variables per function to ten. The Regular PSG edition has a free 30-day trial. After installing PSG, case study projects can be viewed in the Case Studies folder in File tab of the PSG menu. In order to modify an existing case study, it should be copied into Work directory. Tutorials about PSG and case study descriptions can be found in Help tab of the menu.

Acknowledgments

We thank Dr. Bogdan Grechuk at the University of Leicester, UK, and Dr. Anton Molyboha at Teza Technologies for carefully proofreading the portions of the manuscript and for providing valuable comments. We are also grateful to Dr. Viktor Kuzmenko at American Optimal Decisions, Inc. for the technical support and assistance with running numerical experiments for the case studies. We express our deepest gratitude to our families for their constant encouragement, support, and patience.

Hoboken, NJ, USA Gainesville, FL, USA Michael Zabarankin Stan Uryasev

Contents

Part I Selected Concepts of Statistical Decision Theory

1	Ran	ndom Variables		
1.1 Probability Distribution Function			3	
1.2 Moments and Moment-Generating Function		Moments and Moment-Generating Function	5	
		1.2.1 Classical Inequalities with Moments	6	
		1.2.2 Moment-Generating Function	7	
	1.3	Partial Moments and Stochastic Dominance	8	
		1.3.1 Performance Functions	9	
		1.3.2 Stochastic Dominance	9	
	1.4	Quantile Functions and Fenchel Duality	10	
	1.5	Entropy and Distribution Divergence Measures	13	
		1.5.1 Shannon Entropy and Kullback–Leibler Divergence	13	
		1.5.2 Renyi Entropy and Renyi Divergence	15	
2	Devi	ation, Risk, and Error Measures	19	
	2.1	Deviation Measures	21	
	2.2	Risk Envelopes and Risk Identifiers	23	
	2.3	Averse Measures of Risk	26	
	2.4	Error Measures	29	
3	Prob	pabilistic Inequalities	33	
	3.1	Basic Probabilistic Inequalities	33	
	3.2	Chebyshev's Inequalities with Deviation Measures	35	
		3.2.1 One-Sided Chebyshev's Inequalities	35	
		3.2.2 Two-Sided Chebyshev's Inequalities	39	
	3.3	Kolmogorov's Inequalities with Deviation Measures	40	
Pa	rt II	Statistical Decision Problems		
4	Max	imum Likelihood Method	45	
	4.1	Maximum Likelihood Principle	45	
		4.1.1 Parametric Estimation	45	

Contents

		4.1.2	Nonparametric Estimation	48
		4.1.3	Estimation of Probability Distributions of Random	
			Vectors	50
5	Entr	ony Max	cimization	53
	5.1		on Entropy Maximization	53
	5.2		e Entropy Minimization	58
	5.3		Entropy Maximization	59
	5.4		y Maximization with Deviation Measures	
	5.5		y Maximization with Constraints on Average	02
	0.0		ntile for Different Confidence Levels	67
_	Door			71
6			Iodels	73
	6.1		lized Linear Regression and Error Decomposition	
	6.2		Squares Linear Regression	
		6.2.1	Unconstrained Least Squares Linear Regression	74 75
	6.2	6.2.2	Least Squares Linear Regression with Constraints	
	6.3		n Regression	77
		6.3.1	Unconstrained Median Regression	
	6.1	6.3.2	Median Regression with Constraints	
	6.4		le Regression	
		6.4.1 6.4.2	Unconstrained Quantile Regression	
	6.5		Quantile Regression with Constraints	
	6.5	6.5.1	Types of Linear Regression	
		6.5.2	Mixed Quantile Regression	
		6.5.3	Unbiased Linear Regression	
		6.5.4	Risk Acceptable Regression	86
		0.5.4	Robust Regression	
7			L	89
	7.1	_	c Regression	89
	7.2		t Vector Machine	92
		7.2.1	Standard SVM	92
		7.2.2	SVM with Soft Margin	93
		7.2.3	Alternative Formulations of SVM with Soft Margin	95
		7.2.4	Choice of Mapping ϕ : Kernel Approach	97
8	Statis		cision Models with Risk and Deviation	
	8.1		ization of Risk and Deviation	
	8.2		on Minimization in Portfolio Selection	104
		8.2.1	Markowitz Portfolio Selection	104
		8.2.2	Markowitz-Type Portfolio Selection	106
		8.2.3	Generalized Capital Asset Pricing Model	109
		8.2.4	CAPM Reinterpretation: Inverse Portfolio Problem	113
	8.3		on Minimization in Optimal Hedging	115
	8.4		Matching with Deviation Constraints	116
	8.5	Rick M	Inimization in Portfolio Theory	116

Contents xi

	8.6	Applications with Downside Risk Constraints	119
	8.7	Portfolio Optimization with Drawdown Measure	122
		8.7.1 Drawdown Measure	122
		8.7.2 Linear Programming Representations	124
		8.7.3 Portfolio Problem Formulations	125
		8.7.4 CAPM with CDaR	127
Pa	rt III	Portfolio Safeguard Case Studies	
9	Portf	olio Safeguard Case Studies	133
	9.1	Case Study: Optimal Hedging of CDO Book	
		(meanabs_dev, meanabs_pen, polynom_abs, cardn)	133
		9.1.1 Problem 1: problem_cdohedge_1	135
		9.1.2 Problem 2: problem_cdohedge_2	136
	9.2	Hedging Portfolio of Options	137
	9.3	Mortgage Pipeline Hedging	142
		9.3.1 Problem 1: problem_cvar_dev	144
		9.3.2 Problem 2: problem_meanabs_dev	145
		9.3.3 Problem 3: problem_std_dev	146
		9.3.4 Problem 4: problem_two_tailed_var75	147
		9.3.5 Problem 5: problem_two_tailed_var90	147
	9.4	Cash Matching Bond Portfolio	148
	9.5	Cash Flow Matching with CVaR Constraints	152
	9.6	Relative Entropy Minimization	157
	9.7	Portfolio Replication with Risk Constraints	159
	9.8	Style Classification with Quantile Regression	162
	9.9	CVaR Estimation Through Explanatory Factors with	
		Mixed Quantile Regression	165
	9.10	Optimal Test Selection	169
		9.10.1 Problem 1: problem_deterministic_LP_model	170
		9.10.2 Problem 2: problem_robust_model	171
		9.10.3 Problem 3: problem_stochastic_model	172
	9.11	Sparse Signal Reconstruction: A Cardinality Approach	173
		9.11.1 Problem 1: problem_constr_cardinality	175
		9.11.2 Problem 2: problem_minimize_cardinality	176
		9.11.3 Problem 3: problem_constr_polynomabs	178
	9.12	Sparse Reconstruction Problems from SPARCO Toolbox	179
		9.12.1 Problem 1: L1 Relaxed	181
		9.12.2 Problem 2: L1 Relaxed D	182
		9.12.3 Problem 3: L2 D	183
	9.13	Optimal Intensity-Modulated Radiation Therapy	
		Treatment-Planning Problem	184
	9.14	Portfolio Optimization—CVaR Versus Standard Deviation	188
		9.14.1 Problem 1: problem_st_dev_covariances_2p9	190
		9.14.2 Problem 2: problem_st_dev_scenarios_2p9	191
		9.14.3 Problem 3: problem min cvar dev 2p9	192

xii Contents

9.15	Portfoli	o Credit-Risk Optimization Modeled by Scenarios	
	and Mix	ctures of Normal Distributions	193
	9.15.1	Problem 1: problem_var_LLN	197
	9.15.2	Problem 2: problem_cvar_LLN	198
	9.15.3	Problem 3: problem_avg_var_CLT	199
	9.15.4	Problem 4: problem_avg_cvar_CLT	200
	9.15.5	Problem 5: problem_avg_var_CLT (Alternative	
		Formulation)	201
	9.15.6	Problem 6: problem_avg_cvar_CLT	
	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	(Alternative Formulation)	203
9.16	ν-SVM	with VaR and CVaR	204
,,,,	9.16.1	Problem 1a: v-SVM with CVaR	206
	9.16.2	Problem 1b: VaR-SVM	207
	9.16.3	Problem 2a: Extended <i>v</i> -SVM with CVaR	208
	9.16.4	Problem 2b: Extended VaR-SVM	209
	9.16.5	Problem 3a: Robust <i>v</i> -SVM with CVaR	210
	9.16.6	Problem 3b: Robust VaR-SVM	211
9.17		Copula CDO Pricing Model: Entropy Approach	212
7.17	9.17.1	Problem 1a: problem_hull_meansquare	214
	9.17.2	Problem 1b: problem_hull_variance	215
	9.17.3	•	215
9.18		Regression and Regularized Logistic Regression	21.
7.10		mating the Probability of Cesarean Section	217
	9.18.1	Problem 1: Logistic Regression	218
	9.18.2	Problem 2: Regularized Logistic Regression	219
9.19		o Optimization with Drawdown Constraints	21,
7.17		ngle Sample Path	220
	9.19.1	Problem 1: Constraint on Maximum Drawdown	222
	9.19.1	Problem 2: Constraint on Average Drawdown	223
	9.19.2	Problem 3: Constraint on CDaR	224
9.20		o Optimization with Drawdown Constraints	22-
9.20		ciple Sample Paths	225
	9.20.1	Problem 1: Constraint on Maximum Drawdown	228
	9.20.1	Problem 2: Constraint on Average Drawdown	230
	9.20.2	Problem 3: Constraint on CDaR	231
9.21		o Optimization with Drawdown Constraints:	231
9.21		Path Versus Multiple Paths	233
	9.21.1	Problem 1: Multiple Sample Paths	236
	9.21.2	Problem 2: Aggregated Sample Path	238
Deference	100		241
Keierenc	es		241
Indov			245

Chapter 3 Probabilistic Inequalities

In various statistical decision problems dealing with safety and reliability, risk is often interpreted as the probability of a dread event or disaster, and minimizing the probability of a highly undesirable event is known as the *safety-first principle* [50]. If the CDF of X is either unknown or complex, the probability in question can be estimated through more simple characteristics such as mean and standard deviation of X, for example, by Markov's and Chebyshev's inequalities. Also, if the probability depends on decision variables, then, in general, an optimization problem, in which it is either minimized or constrained, is nonconvex. In this case, the probability can be estimated by an appropriate probabilistic inequality, and then the optimization problem can be approximated by a convex one; see, e.g., [3, 32].

3.1 Basic Probabilistic Inequalities

Markov's inequality is one of the basic and yet most important probabilistic inequalities. It is given by

$$\mathbb{P}[|X| \ge a] \le \frac{E[|X|]}{a}, \quad a > 0.$$
 (3.1.1)

Its proof is straightforward:

$$|X| \ge |X| I_{\{|X| \ge a\}} \ge a I_{\{|X| \ge a\}},$$

where $I_{\{|X| \geq a\}}$ is the indicator function equal to 1 if the condition in the curly brackets is true and equal to 0 otherwise. Consequently,

$$E[|X|] \ge E[a \ I_{\{|X| \ge a\}}] \equiv a \ \mathbb{P}[|X| \ge a].$$

Despite its simplicity, (3.1.1) is the major source for obtaining other well-known basic probabilistic inequalities:

• Replacing |X| and a in (3.1.1) by $|X|^p$ and a^p , respectively, where p > 0, we obtain

$$\mathbb{P}[|X| \ge a] \le \left(\frac{\|X\|_p}{a}\right)^p, \qquad a > 0.$$

• For $Y = \ln |X|$ and $b = \ln a$, (3.1.1) is transformed into the inequality for estimating the probability of the right tail $Y \ge b$:

$$\mathbb{P}[Y \geq b] \leq E\left[e^{Y-b}\right].$$

• Substituting $(X - E[X])^2$ for |X| and $a \sigma^2(X)$ for a in (3.1.1), we obtain *Chebyshev's inequality*:

$$\mathbb{P}[|X - \mu| \ge a] \le \frac{\sigma^2(X)}{a^2}, \qquad a > 0,$$

which evaluates the probability of how significantly a random variable X deviates from its expected value $\mu = E[X]$ in terms of the standard deviation $\sigma(X)$.

There is also *one-sided Chebyshev's inequality*, also called *Cantelli's inequality*, that estimates the probability of X either not to exceed a given threshold τ :

$$\mathbb{P}[X \le \tau] \le \frac{\sigma^2(X)}{\sigma^2(X) + (\mu - \tau)^2}, \qquad \tau \le \mu, \tag{3.1.2}$$

or not to drop below the threshold τ :

$$\mathbb{P}[X \ge \tau] \le \frac{\sigma^2(X)}{\sigma^2(X) + (\tau - \mu)^2}, \qquad \tau \ge \mu. \tag{3.1.3}$$

Though (3.1.2) and (3.1.3) do not follow from (3.1.1) as simply as two-sided Chebyshev's inequality, their proof still relies on Markov's inequality. For example, for (3.1.2), consider

$$\mathbb{P}[X \le \tau] = \mathbb{P}[t - X \ge t - \tau] \le \mathbb{P}\left[(t - X)^2 \ge (t - \tau)^2\right] \le \frac{E\left[(t - X)^2\right]}{(t - \tau)^2},$$

where t is an arbitrary real number greater than τ and where the last inequality follows from (3.1.1). Since the inequality $\mathbb{P}[X \leq \tau] \leq E\left[(t-X)^2\right]/(t-\tau)^2$ holds for any $t \geq \tau$, setting the derivative of its right-hand side with respect to t to zero, we obtain that $t^* = \mu + \sigma^2(X)/(\mu - \tau) > \tau$ is the minimizer, and $\mathbb{P}[X \leq \tau] \leq E\left[(t^* - X)^2\right]/(t^* - \tau)^2$ reduces to (3.1.2). Observe that the condition $t^* \geq \tau$ holds if and only if $\mu \geq \tau$.

As an immediate application of (3.1.2), consider a probabilistic constraint (also known as *chance constraint*)

$$\mathbb{P}[X \le \tau] \le \alpha,\tag{3.1.4}$$

where $\alpha \in (0, 1)$ is given. Then one-sided Chebyshev's inequality (3.1.2) implies that (3.1.4) is guaranteed to hold if

$$E[X] - \sigma(X)\sqrt{\alpha^{-1} - 1} \ge \tau, \tag{3.1.5}$$

which is a simple and frequently used condition provided that E[X] and $\sigma(X)$ are either known or easy to estimate.

Also, since $\mathbb{P}[X \leq \tau] \equiv F_X(\tau)$, by integrating (3.1.2) with respect to τ , we obtain an estimate for $F_X^{(2)}(\tau) = \int_{-\infty}^{\tau} F_X(s) ds$:

$$F_X^{(2)}(\tau) \le \sigma(X) \left(\frac{\pi}{2} - \arctan\left(\frac{\mu - \tau}{\sigma(X)} \right) \right), \qquad \tau \le \mu.$$

Markov's and Chebyshev's inequalities are sources for other many remarkable probabilistic relationships and are proved to be invaluable in decision problems with insufficient statistical data; see, e.g., [3, 32].

Another useful and frequently used probabilistic inequality is that of Kolmogorov. Suppose X_1, \ldots, X_n are a sequence of independent random variables such that $E[X_k] = 0$ and $\sigma(X_k) < \infty$, $k = 1, \ldots, n$, and let $S_k = \sum_{j=1}^k X_j$. Then *Kolmogorov's inequality* estimates the probability of $\max_{1 \le k \le n} |S_k|$ to exceed a threshold a in terms of $\sigma(S_n)$:

$$\mathbb{P}\left[\max_{1\leq k\leq n}|S_k|\geq a\right]\leq \frac{1}{a}\sum_{k=1}^n\sigma^2(X_k)\equiv \frac{\sigma^2(S_n)}{a}.$$
 (3.1.6)

Chebyshev's and Kolmogorov's inequalities can be improved if the standard deviation is replaced by another deviation measure. Next sections present generalizations of Chebyshev's and Kolmogorov's inequalities and discuss application of generalized inequalities in statistical decision problems.

3.2 Chebyshev's Inequalities with Deviation Measures

3.2.1 One-Sided Chebyshev's Inequalities

The problem of generalizing one-sided Chebyshev's inequality for *law-invariant* deviation measures, 1 e.g., σ , σ -, MAD, and $\text{CVaR}_{\alpha}^{\Delta}$, is formulated as follows: for

¹A deviation measure $\mathscr{D}(X)$ is law invariant if for any two random variables X_1 and X_2 having the same probability distribution, $\mathscr{D}(X_1) = \mathscr{D}(X_2)$, i.e., if $\mathscr{D}(X)$ depends only on the probability distribution of X.

law-invariant $\mathcal{D}: \mathcal{L}^p(\Omega) \to [0,\infty], 1 \leq p < \infty$, and fixed a > 0, find a function $g_{\mathcal{D}}(d)$ such that

$$\mathbb{P}[X \le \mu - a] \le g_{\mathscr{D}}(\mathscr{D}(X)) \quad \text{for all } X \in \mathscr{L}^p(\Omega), \tag{3.2.1}$$

where $\mu = E[X]$, under the conditions: (i) $g_{\mathscr{D}}$ is independent of the distribution of X and (ii) $g_{\mathscr{D}}$ is the least upper bound in (3.2.1), i.e., for every d > 0, there is a random variable X such that (3.2.1) becomes the equality with $\mathscr{D}(X) = d$.

The inequality (3.2.1) can be reformulated as an optimization problem

$$u_{\mathscr{D}}(\delta) = \inf_{X \in \mathscr{L}^{p}(\Omega)} \mathscr{D}(X)$$
subject to $X \in \mathscr{U} = \{X \mid E[X] = 0, \ \mathbb{P}[X \le -a] \ge \delta\},$

$$(3.2.2)$$

with the function $g_{\mathscr{D}}$ determined by

$$g_{\mathscr{D}}(d) = \sup_{\delta \in (0,1)} \{ \delta \mid u_{\mathscr{D}}(\delta) \le d \}; \tag{3.2.3}$$

see [17] for details. Proposition 3 in [17] proves that (3.2.2) is equivalent to minimizing \mathcal{D} over a subset of \mathcal{U} , whose elements are undominated random variables with respect to *convex ordering*,² and that a solution of (3.2.2) is the random variable $X^*(\delta)$ assuming only two values -a and $\delta a/(1-\delta)$ with the probabilities δ and $1-\delta$, respectively, i.e.,

$$\mathbb{P}[X^*(\delta) = -a] = \delta, \qquad \mathbb{P}\left[X^*(\delta) = \frac{\delta a}{1 - \delta}\right] = 1 - \delta.$$

Thus,

$$u_{\mathscr{D}}(\delta) = \mathscr{D}(X^*(\delta)),$$

and if $u_{\mathscr{D}}$ has the inverse $u_{\mathscr{D}}^{-1}$, then (3.2.3) implies that $g_{\mathscr{D}}(d) = u_{\mathscr{D}}^{-1}(d)$.

For σ , MAD, σ_- , and CVaR $^{\Delta}_{\alpha}$, the function $u_{\mathscr{D}}(\delta)$ and its inverse, which is $g_{\mathscr{D}}(d)$, are given by

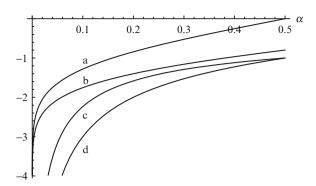
$$u_{\sigma}(\delta) = \sigma(X^*(\delta)) = a\sqrt{\frac{\delta}{1-\delta}}, \qquad g_{\sigma}(d) = u_{\sigma}^{-1}(d) = \frac{d^2}{a^2+d^2},$$

$$u_{\text{MAD}}(\delta) = \text{MAD}(X^*(\delta)) = 2a\delta, \qquad g_{\text{MAD}}(d) = u_{\text{MAD}}^{-1}(d) = \frac{d}{2a},$$

$$u_{\sigma_{-}}(\delta) = \sigma_{-}(X^*(\delta)) = a\sqrt{\delta}, \qquad g_{\sigma_{-}}(d) = u_{\sigma_{-}}^{-1}(d) = \frac{d^2}{a^2},$$

 $^{^2}X$ dominates Y with respect to convex ordering if $E[f(X)] \ge E[f(Y)]$ for any convex function $f: \mathbb{R} \mapsto \mathbb{R}$, which is equivalent to the conditions E[X] = E[Y] and $\int_{-\infty}^x F_X(t)dt \ge \int_{-\infty}^x F_Y(t)dt$ for all $x \in \mathbb{R}$, where F_X and F_Y are CDFs of X and Y, respectively.

Fig. 3.1 Comparison of $q_Z^+(\alpha)$ (curve a) with $\overline{q}_Z(\alpha)$ (curve b), $-\sqrt{\alpha^{-1}-1}$ (curve c), and $-1/\sqrt{2\alpha}$ (curve d) for a standard normal random variable Z for $\alpha \in (0,1/2]$



where $\Phi^{-1}(\alpha)$ is the inverse of the CDF of Z: $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-s^2/2} ds$. Figure 3.1 shows $q_Z^+(\alpha) = \Phi^{-1}(\alpha)$, $-\sqrt{\alpha^{-1}-1}$, $-1/\sqrt{2\alpha}$, and $\overline{q}_Z(\alpha)$ for $\alpha \in (0,1/2]$. For very small α , $\overline{q}_Z(\alpha)$ is close to $q_Z^+(\alpha)$, whereas $-\sqrt{\alpha^{-1}-1}$, which corresponds to σ , is a quite conservative bound over the whole range.

The following general result holds. Let

$$\mathcal{C}_{\sigma} = \left\{ X \mid E[X] - \sigma(X) \sqrt{\alpha^{-1} - 1} \geq \tau \right\}, \qquad \mathcal{C}_{\overline{q}_X} = \left\{ X \mid \overline{q}_X(\alpha) \geq \tau \right\}$$

be the feasible sets of X for the constraints (3.1.5) and (3.2.11), respectively. Then

$$\mathscr{C}_{\sigma} \subseteq \mathscr{C}_{\overline{q}_{Y}};$$

see the proof in [32]. It shows that the constraint (3.2.11) yields a larger feasible set than (3.1.5) does. Also, for a discretely distributed random variable X, $\overline{q}_X(\alpha)$ in (3.2.11) can be reformulated as the linear program (1.4.5), which is attractive from the computational perspective; see [32]. However, the advantage of (3.1.5) is in its simplicity and possibility to obtain closed-form analytical solutions; see [32].

3.2.2 Two-Sided Chebyshev's Inequalities

The problem for generalizing two-sided Chebyshev's inequality for an arbitrary law-invariant deviation measure is formulated similarly to (3.2.1); see [17] for details.

As in (3.2.1), let $\mu = E[X]$ and a > 0. Two-sided Chebyshev's inequality with MAD, σ_- , and $\text{CVaR}_{\alpha}^{\Delta}$ is given by

$$\mathbb{P}[|X - \mu| \ge a] \le \frac{\mathsf{MAD}(X)}{a},\tag{3.2.13}$$

Chapter 5 Entropy Maximization

The previous chapter showed that given independent observations of a random variable X, the probability distribution of X can be estimated based on the maximum likelihood principle. However, if no observations of X are available, but some integral characteristics of the distribution of X are known, for example, mean μ and standard deviation σ , the main principle for finding the distribution in question is, arguably, the one of *maximum entropy*. This principle, also known as *MaxEnt*, originated from the information theory and statistical mechanics (see [22]) and determines the "most unbiased" probability distribution for X subject to any constraints on X (prior information). Nowadays, it is widely used in financial engineering and statistical decision problems [4, 11, 56]. Estimation of probability distributions through entropy maximization and through relative entropy minimization subject to various constraints on unknown distributions is the subject of this chapter.

5.1 Shannon Entropy Maximization

A classical application of the maximum entropy principle in statistics is estimating the probability distribution of a random variable $X \in \mathcal{L}^m(\Omega)$ provided that the first m moments of X are known to be $\mu_1 \in \mathbb{R}, \ldots, \mu_m \in \mathbb{R}$:

$$\max_{X \in \mathscr{L}^m(\Omega)} S(X) \qquad \text{subject to} \quad E[X^k] = \mu_k, \quad k = 1, \dots, m, \tag{5.1.1}$$

where S(X) is the Shannon entropy of X.

If X is restricted to assume only n distinct values $x_1 \in \mathbb{R}, \ldots, x_n \in \mathbb{R}$ with nonnegative probabilities p_1, \ldots, p_n summing to 1, then the problem (5.1.1) takes the form

Example 5.1 (No prior information, m = 0). If no moments of a random variable X are known, then m = 0 and the maximum entropy distributions that solve (5.1.2) and (5.1.3) are uniform:

$$p_1 = \dots = p_n = \frac{1}{n}$$
 in discrete case,
 $f_X(t) = \frac{1}{b-a} I_{\{t \in [a,b]\}}$ in continuous case. (5.1.6)

In other words, the maximum entropy principle implies that without any information about a random variable X (either discretely or continuously distributed), all outcomes of X should be equally probable.

Example 5.2 (Known mean, m=1). If it is only known that the mean of a random variable X is μ , then m=1 and the maximum entropy distributions (5.1.4) and (5.1.5) take the form

$$p_k = \frac{e^{\rho x_k}}{\sum_{k=1}^n e^{\rho x_k}}, \quad k = 1, \dots, n,$$
 (5.1.7)

and

$$f_X(t) = \frac{\lambda e^{\lambda t}}{e^{\lambda b} - e^{\lambda a}} I_{\{t \in [a,b]\}},$$
(5.1.8)

respectively, where ρ satisfies $\sum_{k=1}^{n} (x_k - \mu) e^{\rho x_k} = 0$ and $\lambda > 0$ is found from the equation $(\lambda(\mu - a) + 1)e^{\lambda a} = (\lambda(\mu - b) + 1)e^{\lambda b}$.

Corollary 5.1 (Known mean and semi-infinite support). Example 5.2 implies that

(a) If X is known to assume only integers starting from 1, i.e., $x_k = k$ with k = 1, 2, ... and $\mu > 1$, then $e^{\rho} = 1 - 1/\mu$ and the maximum entropy distribution (5.1.7) reduces to

$$p_k = \frac{(\mu - 1)^{k-1}}{\mu^k}, \quad k = 1, 2, \dots$$

(b) If X is known to be continuously distributed on $[0, \infty)$ with $\mu > 0$, then $\lambda = -1/\mu$ and the maximum entropy solution (5.1.8) simplifies to the exponential distribution $f_X(t) = e^{-t/\mu}/\mu$, $t \in [0, \infty)$.

Example 5.3 (Known mean and standard deviation, m=2). Given that a random variable X has mean μ and variance σ^2 , m=2 and the maximum entropy distribution (5.1.4) reduces to

$$p_k = \frac{\exp\left(\rho_1 x_k + \rho_2 x_k^2\right)}{\sum_{k=1}^n \exp\left(\rho_1 x_k + \rho_2 x_k^2\right)}, \quad k = 1, \dots, n,$$
 (5.1.9)

where ρ_1 and ρ_2 are found from the system

$$\begin{cases} \sum_{k=1}^{n} (x_k - \mu) \exp(\rho_1 x_k + \rho_2 x_k^2) = 0, \\ \sum_{k=1}^{n} (x_k^2 - \mu^2 - \sigma^2) \exp(\rho_1 x_k + \rho_2 x_k^2) = 0, \end{cases}$$

whereas (5.1.5) simplifies to

$$f_X(t) = \frac{\exp(\lambda_1 t + \lambda_2 t^2)}{\int_a^b \exp(\lambda_1 t + \lambda_2 t^2) dt} I_{\{t \in [a,b]\}}$$
 (5.1.10)

with λ_1 and λ_2 found from the system

$$\begin{cases} \int_a^b (t - \mu) \exp\left(\lambda_1 t + \lambda_2 t^2\right) dt = 0, \\ \int_a^b \left(t^2 - \mu^2 - \sigma^2\right) \exp\left(\lambda_1 t + \lambda_2 t^2\right) dt = 0. \end{cases}$$

Unlike Example 5.2, Example 5.3 does not offer simplifications for discretely distributed random variables assuming infinitely many integer values either on \mathbb{R}^+ or \mathbb{R} . However, it yields an important corollary for a random variable continuously distributed on \mathbb{R} .

Corollary 5.2 (Continuously distributed random variable on \mathbb{R} with given mean and variance). If only mean μ and variance σ^2 of a continuously distributed random variable X on \mathbb{R} are known, the maximum entropy PDF is given by $\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-(t-\mu)^2/(2\sigma^2)\right)$. In other words, X is normally distributed with mean μ and variance σ^2 .

A generalization of the maximum entropy problem (5.1.3) is given by

$$\max_{f_X(t)} - \int_V f_X(t) \ln f_X(t) dt$$
subject to
$$\int_V h_k(t) f_X(t) dt = a_k, \quad k = 1, \dots, m,$$

$$\int_V f_X(t) dt = 1, \qquad f_X(t) \ge 0, \quad t \in V, \tag{5.1.11}$$

where V is a given closed support set $V \subseteq \mathbb{R}$ of $f_X(t)$, so that $f_X(t) \equiv 0$ for $t \notin V$; h_1, \ldots, h_m are given measurable functions; and a_1, \ldots, a_m are given constants.

Boltzmann's theorem [10, Theorem 12.1.1] shows that if there exist $\lambda_1, \ldots, \lambda_n$, and c > 0 such that the PDF

$$f_X(t) = c \exp\left(\sum_{k=1}^m \lambda_k h_k(t)\right) I_{\{t \in V\}}$$
 (5.1.12)

satisfies the constraints in (5.1.11), then (5.1.12) is the global maximum of (5.1.11).

With arbitrary constraints on a random variable X, the maximum entropy problem has no closed-form solution regardless of whether X is distributed continuously or discretely. In this case, maximum entropy probability distributions are found by means of numerical optimization.

The next two examples present entropy maximization problems arising in *collateralized debt obligation* (CDO) pricing models [20].

Example 5.4 (Entropy maximization with no-arbitrage constraints). Suppose there are m CDO tranches and there are n scenarios for the hazard rate (λ in a Poisson process modeling default) with unknown probabilities p_1, \ldots, p_n . Let a_{ij} be the expected net payoff of tranche j in hazard rate scenario i. Under the no-arbitrage assumption, the expected net payoff of each CDO tranche over all hazard rate scenarios should be zero: $\sum_{i=1}^n a_{ij} p_i = 0, j = 1, \ldots, m$. If instead of a_{ij} we use the expected net payoffs \underline{a}_{ij} and \overline{a}_{ij} corresponding to ask and bid quotes for tranche j spread, then no-arbitrage constraints are given by $\sum_{i=1}^n \underline{a}_{ij} p_i \leq 0$ and $\sum_{i=1}^n \overline{a}_{ij} p_i \geq 0, j = 1, \ldots, m$, and the problem is to maximize the Shannon entropy with respect to the hazard rate scenario probabilities p_1, \ldots, p_n subject to the no-arbitrage constraints:

$$\max_{p_1,\dots,p_n} -\sum_{i=1}^n p_i \ln p_i$$
subject to
$$\sum_{i=1}^n p_i = 1, \quad p_i \ge 0, \quad i = 1,\dots,n,$$

$$\sum_{i=1}^n \underline{a}_{ij} p_i \le 0, \quad j = 1,\dots,m,$$

$$\sum_{i=1}^n \overline{a}_{ij} p_i \ge 0, \quad j = 1,\dots,m.$$
no-arbitrage constraints
$$\sum_{i=1}^n \overline{a}_{ij} p_i \ge 0, \quad j = 1,\dots,m.$$

Example 5.5 (Entropy maximization with no-arbitrage constraints and constraints on distribution shape). This problem is similar to (5.1.13). It imposes additional constraints on the distribution to have a bell shape (hump):

$$\max_{p_1,\dots,p_n} -\sum_{i=1}^n p_i \ln p_i$$
subject to
$$\sum_{i=1}^n p_i = 1, \quad p_i \ge 0, \quad i = 1,\dots,n,$$

$$\sum_{i=1}^n \underline{a}_{ij} \, p_i \le 0, \quad j = 1,\dots,m,$$

$$\sum_{i=1}^n \overline{a}_{ij} \, p_i \ge 0, \quad j = 1,\dots,m,$$

$$1 \le w_l \le w_r \le n,$$

$$\frac{p_{i-1} + p_{i+1}}{2} \ge p_i, \quad i = 2,\dots,w_l - 1 \qquad \text{(left slope is convex)}$$

$$\frac{p_{i-1} + p_{i+1}}{2} \le p_i, \quad i = w_l + 1,\dots,w_r - 1 \quad \text{(hump is concave)}$$

$$\frac{p_{i-1} + p_{i+1}}{2} \ge p_i, \quad i = w_r + 1,\dots,n - 1 \quad \text{(right slope is convex)}$$

$$(5.1.14)$$

where w_l and w_r are indices of points of inflection.

In Sect. 9.17, the case study "Implied Copula CDO Pricing Model: Entropy Approach" implements the problems (5.1.13) and (5.1.14) in Portfolio Safeguard and solves the problems with real-life data.

5.2 Relative Entropy Minimization

The problem (4.1.4) provides an important insight: maximizing the log-likelihood function in (4.1.4) is equivalent to minimizing the *relative entropy* or *Kullback–Leibler divergence measure*:

$$D_{KL}(Y||X) = \sum_{k=1}^{n} q_k \ln \frac{q_k}{p_k} = \sum_{k=1}^{n} q_k \ln q_k - \sum_{k=1}^{n} q_k \ln p_k$$

with respect to unknown distribution $p = (p_1, ..., p_l)$ of X given the sample distribution $q = (q_1, ..., q_l)$ of Y, where $q_k = n_k/n$, k = 1, ..., l.

This observation has far-reaching implications: q should not necessarily be a sample distribution and can be replaced either by a prior probability distribution or by an arbitrary reference probability distribution.

However, $D_{KL}(Y||X) \neq D_{KL}(X||Y)$, and the minimum relative entropy principle (MinEnt) or the principle of minimum discrimination information aims to find a random variable X that minimizes the relative entropy $D_{KL}(X||Y)$ for a given reference random variable Y subject to any additional constraints on X:

$$\min_{X \in \mathcal{X}} D_{\text{KL}}(X||Y)$$
 where \mathcal{X} is a feasible set of X .

It is widely used in statistical decision problems dealing with estimation of unknown probability distributions under various constraints. If Y is uniformly distributed, then minimizing $D_{\mathrm{KL}}(X||Y)$ is equivalent to maximizing the Shannon entropy S(X).

Example 5.6 (Relative entropy minimization with linear constraints). The problem of finding a discrete probability distribution $p = (p_1, \ldots, p_n)$ closest to a given probability distribution $q = (q_1, \ldots, q_n)$ in the sense of relative entropy subject to linear constraints on p is formulated by

$$\min_{p_1,\dots,p_n} \sum_{i=1}^n p_i \ln \frac{p_i}{q_i}$$
subject to
$$\sum_{i=1}^n p_i = 1, \quad Ap \le b,$$

$$l_i \le p_i \le u_i, \quad i = 1,\dots,n,$$
(5.2.1)

where real-valued matrix $A = \{a_{ij}\}_{i,j=1}^{n,m}$ and vector $b \in \mathbb{R}^m$ are known and l_i and u_i are lower and upper bounds such that $0 \le l_i \le u_i \le 1, i = 1, \dots, n$.

5.3 Renyi Entropy Maximization

The Renyi entropy (1.5.4) and (1.5.5) can be used in place of the Shannon entropy (1.5.1) and (1.5.2) in entropy maximization problems. For $\alpha \neq 1$, maximizing the Renyi entropy is equivalent to maximizing $\frac{1}{1-\alpha}\sum_{k=1}^n p_k^{\alpha}$ in the discrete case and to maximizing $\frac{1}{1-\alpha}\int_a^b f_X(t)^{\alpha}dt$ in the continuous case. Thus, the entropy maximization problem (5.1.2) with the Renyi entropy for $\alpha \neq 1$ ($\alpha > 0$) is formulated by

5.4 Entropy Maximization with Deviation Measures

How to estimate the probability distribution of a random variable X if its mean and a deviation measure other than the standard deviation are known?

The problem of maximizing the Shannon entropy S(X) for a continuously distributed random variable $X \in \mathcal{L}^1(\Omega)$, whose mean and law-invariant deviation $\mathscr{D}: \mathcal{L}^p(\Omega) \mapsto [0,\infty], \ p \in [1,\infty]$, are known to be μ and d, respectively, is formulated by

$$\max_{X \in \mathcal{L}^1(\Omega)} S(X) \qquad \text{subject to} \quad E[X] = \mu, \quad \mathscr{D}(X) = d. \tag{5.4.1}$$

Let $X_0 \in \mathcal{L}^1(\Omega)$ be a new random variable with a PDF $f_{X_0}(t)$, and let $X = d X_0 + \mu$. Then the PDF and the entropy of X are given by

$$f_X(t) = \frac{1}{d} f_{X_0}\left(\frac{t-\mu}{d}\right), \qquad S(X) = S(X_0) + \ln d,$$

respectively, and the problem (5.4.1) simplifies to

$$\max_{X_0 \in \mathcal{L}^1(\Omega)} S(X_0) \qquad \text{subject to} \quad E[X_0] = 0, \quad \mathcal{D}(X_0) = 1. \tag{5.4.2}$$

For standard deviation, mean absolute deviation (MAD), lower range deviation, standard lower semideviation, and CVaR deviation, the problem (5.4.2) can be recast in the form (5.1.11), and in these cases, solutions to (5.4.2) are given by (5.1.12):

$\mathscr{D}(X_0)$	Support	$f_{X_0}(t)$	$S(X_0)$
$\sigma(X_0)$	$(-\infty, \infty)$	$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right)$	$\frac{1}{2}(1 + \ln[2\pi])$
$MAD(X_0)$	$(-\infty, \infty)$	$\frac{1}{2}\exp(- t)$	$1 + \ln 2$
$E[X_0] - \inf X_0$	$[-1,\infty)$	$\exp(-t-1)$	1
$\sigma_{-}(X_0)$	$(-\infty, \infty)$	$c_1 \exp\left(c_2 t - \frac{1}{2}[-t]_+^2\right)$	1.84434
$\text{CVaR}^{\Delta}_{\alpha}(X_0)$	$(-\infty,\infty)$	$(1-\alpha)\exp\left(c_{\alpha}-t-\frac{1}{\alpha}\left[c_{\alpha}-t\right]_{+}\right)$	$1 - \ln(1 - \alpha)$

where $c_1 \approx 0.260713$, $c_2 \approx -0.638833$, and $c_\alpha = (2\alpha - 1)/(1 - \alpha)$; see [16] for details. Figure 5.1 illustrates the function $f_{X_0}(t)$ for $\text{CVaR}_{\alpha}^{\Delta}$ with $\alpha = 0.01, 0.3, 0.5, 0.7, 0.8$, and 0.9; see Example 5.11 for obtaining $f_{X_0}(t)$.

A problem closely related to (5.4.1) is maximizing the Shannon entropy for a continuously distributed random variable $Y \in \mathcal{L}^1(\Omega)$ subject to a constraint on the deviation \mathcal{D} projected from a nondegenerate error measure \mathcal{E} and subject to a constraint on the statistic \mathcal{L} associated with \mathcal{E} :

$$\max_{Y \in \mathcal{L}^1(\Omega)} S(Y) \quad \text{subject to} \quad \mathcal{D}(Y) = d, \quad c \in \mathcal{S}(Y), \tag{5.4.3}$$

where \mathcal{D} and \mathcal{S} are defined by (2.4.3) and (2.4.4), respectively.

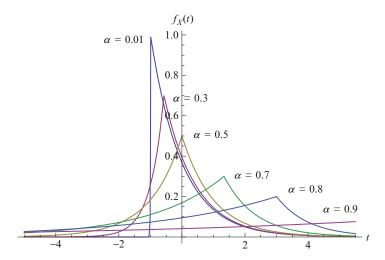


Fig. 5.1 The PDF $f_X(t)$ that maximizes the Shannon entropy S(X) subject to constraints on the mean and CVaR deviation: E[X]=0 and $CVaR^{\Delta}_{\alpha}(X)=1$ for $\alpha=0.01, 0.3, 0.5, 0.7, 0.8$, and 0.9

Let $Y_0 \in \mathcal{L}^1(\Omega)$ be a new random variable with a PDF $f_{Y_0}(t)$, and let $Y = d Y_0 + c$. Then the PDF and the entropy of Y are given by

$$f_Y(t) = \frac{1}{d} f_{Y_0}\left(\frac{t-c}{d}\right), \qquad S(Y) = S(Y_0) + \ln d,$$

respectively, and the problem (5.4.3) simplifies to

$$\max_{Y_0 \in \mathcal{L}^1(\Omega)} S(Y_0) \qquad \text{subject to} \quad \mathcal{D}(Y_0) = 1, \quad 0 \in \mathcal{S}(Y_0). \tag{5.4.4}$$

Proposition 5.1. Let $Z^* \in \mathcal{L}^1(\Omega)$ be a continuously distributed random variable that maximizes the Shannon entropy subject to a constraint on a nondegenerate error measure \mathcal{E} :

$$\max_{Z \in \mathcal{L}^1(\Omega)} S(Z) \quad \text{subject to} \quad \mathcal{E}(Z) = 1. \tag{5.4.5}$$

If the deviations \mathcal{D} in (5.4.1) and in (5.4.3) are projected from \mathcal{E} , then the random variables X_0^* and Y_0^* that solve the problems (5.4.2) and (5.4.4), respectively, are determined by

$$X_0^* = Z^* - E[Z^*], Y_0^* = Z^* - C^*, C^* \in \mathcal{S}(Z^*), (5.4.6)$$

where ${\mathcal S}$ is the statistic associated with ${\mathcal E}$ and

$$-E[Y_0^*] \in \mathscr{S}(X_0^*).$$

Proof. Let X_0 and Y_0 be feasible random variables in (5.4.2) and (5.4.4), respectively, and let $C_0 \in \mathcal{S}(X_0)$. Then (5.4.2) reduces to (5.4.4) by substitution $Y_0 = X_0 - C_0$. Indeed, $S(X_0) = S(Y_0)$, $0 \in \mathcal{S}(Y_0)$, and $\mathcal{D}(X_0) = \mathcal{D}(Y_0)$.

Now, since \mathscr{D} is projected from \mathscr{E} , the constraint $0 \in \mathscr{S}(Y_0)$ implies that $\mathscr{D}(Y_0) = \mathscr{E}(Y_0 - 0) = \mathscr{E}(Y_0)$ and the problem (5.4.4) can be equivalently restated as

$$\max_{Y_0 \in \mathscr{L}^1(\Omega)} S(Y_0) \qquad \text{subject to} \quad \mathscr{E}(Y_0) = 1, \quad 0 \in \mathscr{S}(Y_0). \tag{5.4.7}$$

Let Z^* be a solution to (5.4.5), let $C^* \in \mathcal{S}(Z^*)$, and let $Y_0^* = Z^* - C^*$, then $\mathcal{E}(Y_0^*) = 1$. Indeed, by definition of the statistic \mathcal{S} associated with \mathcal{E} ,

$$\mathscr{E}(Y_0^*) = \mathscr{E}(Z^* - C^*) = \min_{C \in \mathbb{R}} \mathscr{E}(Z^* - C) \le \mathscr{E}(Z^*) = 1,$$

so that $\mathscr{E}(Y_0^*) \leq 1$. By contradiction, let $\mathscr{E}(Y_0^*) = \delta < 1$, and let $\tilde{Z} = Y_0^*/\delta$. Then, the positive homogeneity of \mathscr{E} implies that $\mathscr{E}(\tilde{Z}) = 1$ and $S(\tilde{Z}) = S(Y_0^*) - \ln \delta = S(Z^*) - \ln \delta > S(Z^*)$, so that Z^* is not optimal for (5.4.5). Thus, $\mathscr{E}(Y_0^*) = 1$.

Now, compared to (5.4.5), the problem (5.4.7) has the additional constraint $0 \in \mathcal{S}(Y_0)$, and consequently, its optimal value is less than or equal to that of (5.4.5), i.e., $S(Y_0) \leq S(Z^*)$ for any feasible Y_0 . However, $0 \in \mathcal{S}(Y_0^*)$ and $\mathcal{E}(Y_0^*) = 1$, so that Y_0^* is feasible for (5.4.7), and since $S(Y_0^*) = S(Z^*)$, Y_0^* is also optimal for (5.4.7).

Finally, the constraint $E[X_0]=0$ and the relationship $Y_0=X_0-C_0$ with $C_0\in \mathscr{S}(X_0)$ imply that $E[Y_0]=E[X_0]-C_0=-C_0$. Consequently, $-E[Y_0]\in \mathscr{S}(X_0)$ and $X_0=Y_0+C_0=Y_0-E[Y_0]$, and optimal X_0^* in (5.4.2) is determined by $X_0^*=Y_0^*-E[Y_0^*]=Z^*-C^*-E[Z^*-C^*]=Z^*-E[Z^*]$. \square

For the error measure (2.4.1): $\mathscr{E}(Z) = \|a Z_+ + b Z_-\|_p$ with a > 0, b > 0, and $p \in [1, \infty)$, the problem (5.4.5) can be represented in the form (5.1.11) with m = 1, $V = (-\infty, \infty)$, and

$$h_1(t) = (a \max\{t, 0\} + b \max\{-t, 0\})^p$$
.

Its solution is given by (5.1.12), where c and λ_1 are found from the constraints $\int_{-\infty}^{\infty} f_Z(t) dt = 1$ and $\int_{-\infty}^{\infty} h_1(t) f_Z(t) dt = 1$, so that the PDF of optimal Z^* in (5.4.5) is determined by

$$f_{Z^*}(t) = \frac{\exp\left(-\frac{1}{p}(a\,\max\{t,0\} + b\,\max\{-t,0\})^p\right)}{(a^{-p} + b^{-p})\,p^{1/p}\Gamma[(p+1)/p]}, \qquad t \in \mathbb{R}, \tag{5.4.8}$$

and

$$S(Z^*) = \frac{1 + \ln p}{p} + \ln \left((a^{-p} + b^{-p}) \Gamma \left\lceil \frac{p+1}{p} \right\rceil \right).$$

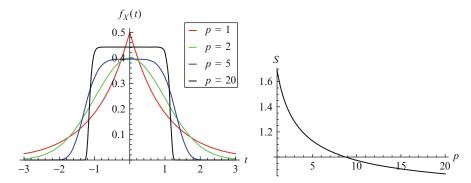


Fig. 5.2 PDFs $f_Z(t)$ that maximize the Shannon entropy S(Z) subject to the constraint $||Z||_p = 1$ for p = 1, 2, 5, 20 and the value of the entropy S(Z) for these PDFs as a function of $p \in [1, \infty)$. The PDF with the sharp spike corresponds to p = 1, and the PDF dispersion decreases with p

Example 5.10 (Entropy maximization with a constraint on p-norm). For a = b = 1, the error measure (2.4.1) simplifies to $||X||_p$, and consequently, the PDF of a random variable Z^* that solves the problem (5.4.5) with $\mathscr{E}(Z) = ||Z||_p$ is given by (5.4.8) with a = b = 1:

$$f_{Z^*}(t) = \frac{1}{2p^{1/p}\Gamma[(p+1)/p]} \exp\left(-\frac{|t|^p}{p}\right), \quad t \in \mathbb{R}.$$
 (5.4.9)

Figure 5.2 shows the function (5.4.9) for p=1,2,5, and 20 and also depicts $S(Z^*)=(1+\ln p)/p+\ln (2\Gamma[(p+1)/p])$ as a function of $p\in[1,\infty)$. Remarkably, $\lim_{n\to\infty} f_{Z^*}(t)=\frac{1}{2}I_{t-1}(t)$ and $\lim_{n\to\infty} S(Z^*)=\ln 2$.

Remarkably, $\lim_{p\to\infty} f_{Z^*}(t) = \frac{1}{2} I_{\{-1 \le t \le 1\}}$ and $\lim_{p\to\infty} S(Z^*) = \ln 2$. In this case, $E[Z^*] = 0$, and (5.4.6) implies that X_0^* solving (5.4.2) is given by $X_0^* = Z^*$, so that $f_{X_0^*}(t) = f_{Z^*}(t)$. Since σ and MAD are projected from the error measures $\|\cdot\|_2$ and $\|\cdot\|_1$, respectively, the PDFs of optimal X_0^* in (5.4.2) with $\mathscr{D} = \sigma$ and $\mathscr{D} = \mathrm{MAD}$ are given by (5.4.9) for p = 2 and p = 1, respectively.

Example 5.11 (Entropy maximization with a constraint on asymmetric mean absolute error). For p=1, a=1, and $b=1/\alpha-1$, the error measure (2.4.1) reduces to (2.0.3), and consequently, the PDF of a random variable Z^* that solves the problem (5.4.5) with $\mathcal{E}(Z) = \mathcal{E}_{\alpha}(Z)$ is given by (5.4.8) with p=1, a=1, and $b=1/\alpha-1$:

$$f_{Z^*}(t) = (1 - \alpha) \exp\left(\frac{1}{\alpha} \min\{t, 0\} - t\right), \qquad t \in \mathbb{R}, \tag{5.4.10}$$

for which

$$E[Z^*] = -\frac{2\alpha - 1}{1 - \alpha}, \qquad q_{Z^*}^+(\alpha) = 0, \qquad S(Z^*) = 1 - \ln(1 - \alpha).$$

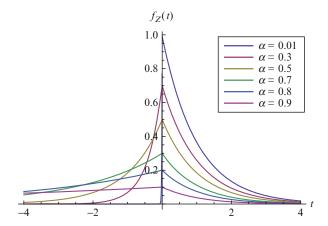


Fig. 5.3 The PDF $f_Z(t)$ that maximizes the Shannon entropy S(Z) subject to a constraint on the asymmetric mean absolute error (2.0.3): $\mathscr{E}_{\alpha}(Z)=1$ for $\alpha=0.01,0.3,0.5,0.7,0.8$, and 0.9

Figure 5.3 illustrates the function $f_{Z^*}(t)$ for $\alpha=0.01,0.3,0.5,0.7,0.8$, and 0.9. The deviation \mathscr{D} projected from \mathscr{E}_{α} and the statistic \mathscr{S} associated with \mathscr{E}_{α} are CVaR deviation and α -quantile, respectively. Let X_0^* solve (5.4.2) with $\mathscr{D}=\mathrm{CVaR}_{\alpha}^{\alpha}$:

$$\max_{X_0 \in \mathscr{L}^1(\Omega)} S(X_0) \qquad \text{subject to} \quad E[X_0] = 0, \quad \text{CVaR}_\alpha^\Delta(X_0) = 1,$$

and let Y_0^* solve (5.4.4) with $\mathscr{D}=\mathrm{CVaR}_\alpha^\Delta$ and $\mathscr{S}(Y_0)=q_{Y_0}(\alpha)$:

$$\max_{Y_0 \in \mathscr{L}^1(\Omega)} S(Y_0) \qquad \text{subject to} \quad \mathrm{CVaR}_\alpha^\Delta(Y_0) = 1, \quad q_{Y_0}^+(\alpha) = 0.$$

Then the relationships (5.4.6) imply that

$$X_0^* = Z^* + c_\alpha, \qquad Y_0^* = Z^*,$$

where $c_{\alpha} = (2\alpha - 1)/(1 - \alpha)$, and the PDF of X_0^* is determined by

$$f_{X_0^*}(t) = (1 - \alpha) \exp\left(\frac{1}{\alpha} \min\{t - c_\alpha, 0\} - t + c_\alpha\right), \qquad t \in \mathbb{R}$$
 (5.4.11)

(see Fig. 5.1), whereas the PDF of Y_0^* coincides with (5.4.10) (see Fig. 5.3).

Another approach to entropy maximization is based on a quantile representation of the entropy. For a continuously distributed X with a CDF $F_X(t)$, the Shannon entropy S(X) can be represented by

References

- [1] Acerbi, C.: Spectral measures of risk: a coherent representation of subjective risk aversion. J. Bank. Financ. **26**(7), 1487–1503 (2002)
- [2] Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Math. Financ. 9(3), 203–228 (1999)
- [3] Bonami, P., Lejeune, M.A.: An exact solution approach for portfolio optimization problems under stochastic and integer constraints. Oper. Res. **57**(3), 650–670 (2009)
- [4] Buckley, J.J.: Entropy principles in decision making under risk. Risk Anal. 5(4), 303–313 (1979)
- [5] Chang, C.C., Lin, C.J.: Training ν-support vector classifiers: theory and algorithms. Neural Comput. **13**, 2119–2147 (2001)
- [6] Chekhlov, A., Uryasev, S., Zabarankin, M.: Portfolio Optimization with Drawdown Constraints, pp. 263–278. Risk Books, London (2003)
- [7] Chekhlov, A., Uryasev, S., Zabarankin, M.: Drawdown measure in portfolio optimization. Int. J. Theor. Appl. Financ. 8(1), 13–58 (2005)
- [8] Cortes, C., Vapnik, V.: Support vector networks. Mach. Learn. 20, 273-297 (1995)
- [9] Costa, J., Hero, A., Vignat, C.: On Solutions to Multivariate Maximum-entropy Problems, vol. 2683, pp. 211–228. Springer, Berlin (2003)
- [10] Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley, New York
- [11] Cozzolino, J.M., Zahner, M.J.: The maximum-entropy distribution of the future market price of a stock. Oper. Res. 21(6), 1200–1211 (1973)
- [12] Crisp, D.J., Burges, C.J.C.: A geometric interpretation of ν-SVM classifiers. Neural Inf. Process. Syst. 12, 244–250 (2000)
- [13] Fölmer, H., Schied, A.: Stochastic Finance, 2nd edn. Walter de Gruyter GmbH & Co., Berlin (2004)
- [14] Grauer, R.R.: Introduction to asset pricing theory and tests. In: Roll, R. (ed.) The International Library of Critical Writings in Financial Economics. Edward Elgar Publishing Inc., Cheltenham (2001)
- [15] Grechuk, B., Zabarankin, M.: Inverse portfolio problem with mean-deviation model. Eur. J. Oper. Res. (2013, to appear)
- [16] Grechuk, B., Molyboha, A., Zabarankin, M.: Maximum entropy principle with general deviation measures. Math. Oper. Res. 34(2), 445–467 (2009)
- [17] Grechuk, B., Molyboha, A., Zabarankin, M.: Chebyshev's inequalities with law invariant deviation measures. Probab. Eng. Informational Sci. 24, 145–170 (2010)
- [18] Hardy, G.E., Littlewood, J.E., Pólya, G.: Inequalities, 2nd edn. Cambridge University Press, New York (1952)

242 References

[19] Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn. Springer, New York (2008)

- [20] Hull, J.C., White, A.D.: Valuing credit derivatives using an implied copula approach. J. Derivatives 14(2), 8–28 (2006)
- [21] Iscoe, I., Kreinin, A., Mausser, H., Romanko, A.: Portfolio credit-risk optimization. J. Bank. Financ. 36(6), 1604–1615 (2012)
- [22] Jaynes, E.T.: Information theory and statistical mechanics. Phys. Rev. 106(4), 620–630 (1957)
- [23] Jensen, J.L.: Surles fonctions convexes et les inégalités entre les valeurs moyennes. Acta Math. **30**(1), 175–193 (1906)
- [24] Johnson, O., Vignat, C.: Some results concerning maximum Rényi entropy distributions. Annales de l'Institut Henri Poincare (B) Probab. Stat. 43(3), 339–351 (2007)
- [25] Kalinchenko, K., Uryasev, S., Rockafellar, R.T.: Calibrating risk preferences with generalized CAPM based on mixed CVaR deviation. J. Risk 15(1), 45–70 (2012)
- [26] Koenker, R., Bassett, G.: Regression quantiles. Econometrica 46, 33–50 (1978)
- [27] Kurdila, A., Zabarankin, M.: Convex Functional Analysis. Birkhauser, Switzerland (2005)
- [28] Levy, H.: Stochastic dominance and expected utility: survey and analysis. Manag. Sci. 38(4), 555–593 (1992)
- [29] Markowitz, H.M.: Portfolio selection. J. Financ. 7(1), 77–91 (1952)
- [30] Markowitz, H.M.: Foundations of portfolio theory. J. Financ. 46, 469–477 (1991)
- [31] Mercer, J.: Functions of positive and negative type, and their connection with the theory of integral equations. Philos. Trans. Roy. Soc. London 209(441–458), 415–446 (1909)
- [32] Molyboha, A., Zabarankin, M.: Stochastic optimization of sensor placement for diver detection. Oper. Res. **60**(2), 292–312 (2012)
- [33] Ogryczak, W., Ruszczyński, A.: On consistency of stochastic dominance and meansemideviation models. Math. Program. 89, 217–232 (2001)
- [34] Ogryczak, W., Ruszczyński, A.: Dual stochastic dominance and related mean-risk models. SIAM J. Optim. 13(1), 60–78 (2002)
- [35] Perez-Cruz, F., Weston, J., Hermann, D.J.L., Schölkopf, B.: Extension of the ν-SVM range for classification. Adv. Learn. Theory Method. Models Appl. **190**, 179–196 (2003)
- [36] Rockafellar, R.T.: Convex Analysis, Princeton Mathematics Series, vol. 28. Princeton University Press, Princeton (1970)
- [37] Rockafellar, R.T.: Coherent approaches to risk in optimization under uncertainty. In: Gray, P. (ed.) Tutorials in Operations Research, pp. 38–61. INFORMS, Hanover (2007)
- [38] Rockafellar, R.T., Royset, J.O.: On buffered failure probability in design and optimization of structures. Reliab. Eng. Syst. Saf. 95, 499–510 (2011)
- [39] Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk 2, 21–41 (2000)
- [40] Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. J. Bank. Financ. 26(7), 1443–1471 (2002)
- [41] Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Deviation measures in risk analysis and optimization. Technical Report 2002–7. ISE Department, University of Florida, Gainesville, FL (2002)
- [42] Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Generalized deviations in risk analysis. Financ. Stoch. **10**(1), 51–74 (2006)
- [43] Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Master funds in portfolio analysis with general deviation measures. J. Bank. Financ. 30(2), 743–778 (2006)
- [44] Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Optimality conditions in portfolio analysis with general deviation measures. Math. Program. **108**(2–3), 515–540 (2006)
- [45] Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Equilibrium with investors using a diversity of deviation measures. J. Bank. Financ. 31(11), 3251–3268 (2007)
- [46] Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Risk tuning with generalized linear regression. Math. Oper. Res. 33(3), 712–729 (2008)
- [47] Roell, A.: Risk aversion in Quiggin and Yaari's rank-order model of choice under uncertainty. Econ. J. **97**(Issue Supplement: Conference papers), 143–159 (1987)

References 243

[48] Rousseeuw, P.J., Driessen, K.: Computing LTS regression for large data sets. Data Min. Knowl. Discov. 12(1), 29–45 (2006)

- [49] Rousseeuw, P., Leroy, A.: Robust Regression and Outlier Detection. Wiley, New York (1987)
- [50] Roy, A.D.: Safety first and the holding of assets. Econometrica 20(3), 431–449 (1952)
- [51] Ruszczyński, A.: Nonlinear Optimization. Princeton University Press, Princeton (2006)
- [52] Schölkopf, B., Smola, A., Williamson, R., Bartlett, P.: New support vector algorithms. Neural Comput. 12, 1207–1245 (2000)
- [53] Sharpe, W.F.: Capital asset prices: a theory of market equilibrium under conditions of risk. J. Financ. 19, 425–442 (1964)
- [54] Sharpe, W.F.: Capital asset prices with and without negative holdings. J. Financ. 46, 489–509 (1991)
- [55] Takeda, A., Sugiyama, M.: ν-support vector machine as conditional value-at-risk minimization. In: Proceedings of the 25th International Conference on Machine Learning (ICML 2008), pp. 1056–1063. Morgan Kaufmann, Montreal, Canada (2008)
- [56] Thomas, M.U.: A generalized maximum entropy principle. Oper. Res. 27(6), 1188–1196 (1979)
- [57] Tobin, J.: Liquidity preference as behavior towards risk. Rev. Econ. Stud. 25(2), 65–86 (1958)
- [58] Tsyurmasto, P., Zabarankin, M., Uryasev, S.: Value-at-risk support vector machine: stability to outliers. J. Comb. Optim. (2014, to appear)
- [59] Venables, W., Ripley, B.: Modern Applied Statistics with S-PLUS, 4th edn. Springer, New York (2002)
- [60] van der Waerden, B.: Mathematische Statistik. Springer, Berlin (1957)
- [61] Wets, R.J.B.: Statistical estimation from an optimization viewpoint. Ann. Oper. Res. 85, 79–101 (1999)
- [62] Yaari, M.E.: The dual theory of choice under risk. Econometrica 55(1), 95–115 (1987)
- [63] Zabarankin, M., Pavlikov, K., Uryasev, S.: Capital asset pricing model (CAPM) with drawdown measure. Eur. J. Oper. Res. (2013, to appear)

Symbols α -quantile, 10, 19 average, 11, 37, 96 lower, 10 upper, 10, 96	cone normal, 102 recession, 103 constraint budget, 190–192, 197–200, 202, 203 chance, 35 expected rate of return, 190–192, 197–200,
A	202, 203
acceptance set, 28	no-arbitrage, 57, 212 no-shorting, 109 self-financing, 116, 150
В	convex ordering, 36
Bayesian statistics, 15	copula, 5
beta	Gaussian, 5
CDaR, 127, 129	credit index, 133
CVaR, 110	CVaR, 11, 19, 28, 85, 117, 120, 138, 143, 152,
MaxDD, 128, 129	159, 194, 198, 200, 203, 222, 228,
standard deviation, 110 standard lower semideviation, 110	235, 236 for loss, 20, 120, 140, 154, 161, 190, 195,
breakdown point, 86	206
buffered probability, 38	maximum, 206
curretta procuernty, so	mixed, 27
	worst-case, 27
C	
CAPM, 105, 109, 127	
CAPM-like relations, 110, 113	D
CDaR, 123	data
multiple sample paths, 225, 227, 231, 233,	linearly separable, 92
235, 236	deviation, 19
single sample path, 220, 222, 224, 233, 236, 238	\mathcal{L}^p -type, 21 comonotone, 67
CDF, 3	CVaR, 20, 21, 25, 27, 37, 39, 40, 82, 109,
joint, 4	115, 142, 144, 188, 189, 192
CDO, 57, 115, 133, 212	max, 22, 40
CDS, 115, 133, 212	mixed, 23, 27, 30, 31, 113
CDS spread, 133	finite, 107

deviation (cont.) law-invariant, 20, 35, 39, 67	asymmetric mean absolute, 20, 29, 31, 73, 83, 165, 167
lower range, 111	inverse projection, 31
lower range dominated, 21, 24, 27	mean square, 76
lower worst-case, 21, 25 mean absolute, 6, 10, 19, 21, 25, 37, 39, 40,	mixed quantile, 31, 83, 167 nondegenerate, 29, 31, 73
115, 133, 135, 142, 144, 145	error decomposition, 73
projected from error, 29, 62	error projection, 29
semi- \mathcal{L}^p type, 21	estimation
standard, 10, 21, 24, 115, 142, 143, 146,	mean, 46, 47
189	probability distribution, 50, 53
upper worst-case, 21	variance, 47
VaR, 109	estimator
two-tailed, 87, 115, 142, 144, 147	consistent, 47
distribution log-concave, 67	maximum likelihood, 46 robust, 86
marginal, 51	unbiased, 47
normal, 4, 7, 46, 47, 118	best linear, 72
multivariate, 5, 50	expected shortfall, 11
standard, 109	expected utility theory, 9
Poisson, 46	
posterior, 15	
prior, 15	F
uniform, 4, 7, 47, 55	Fenchel duality, 12 function
distribution mixture, 15, 118, 193 divergence	average loss, 149
λ , 15	cardinality, 135, 175, 176
Kullback-Leibler, 14, 16, 58, 157	concave, 7
probability, 14	convex, 7, 14
Renyi, 16	CVaR component positive, 169
drawdown, 122	maximum deviation, 150
average, 123	mean absolute penalty, 174, 178, 181, 182
multiple sample paths, 226, 227, 230 single sample path, 220, 222, 223	mean square penalty, 181, 183 moment generating, 7, 8
maximum, 123	partial moment penalty for gain, 164, 167
multiple sample paths, 226–228	partial moment penalty for loss, 164, 167
single sample path, 220, 222	polynomial absolute, 175, 178, 181, 182
dual characterization, 23	
	_
E	G Gini index, 23
entropy	Gilli fildex, 23
collision, 16, 61	
Gibbs, 13	Н
Hartley, 16	hazard rate, 57
min, 16	
relative, 14, 58, 157	
Renyi, 15, 59	I
differential, 16	inequality Cantelli's, 34
Shannon, 13, 16, 53, 212, 213, 215 differential, 13	Chebyshev's, 34
error, 19	one-sided, 34
\mathcal{L}_1 , 85, 133, 135, 159, 161, 173, 174, 179,	Gibbs', 14
181	Hardy-Littlewood's, 10

Jensen's, 7	P
Kolmogorov's, 35, 40	PDF, 3
Lyapunov's, 6, 7	marginal, 4
Markov's, 33	portfolio theory, 19
Schwarz's, 6	principle
infimum	maximum entropy, 17, 20, 53
A-effective, 28	maximum likelihood, 45, 71
essential, 3	maximum log-likelihood, 46
information gain, 15	minimum discrimination information, 15
intercept, 71	minimum relative entropy, 15, 17, 59, 157
	safety first, 33
	problem
K	assignment, 170
kernel, 97	cash flow matching, 116, 120, 148, 152,
neural network, 97	154
polynomial, 97	classification, 162, 204
RBF, 97	return-based style, 82, 166
kurtosis, 5	credit risk, 193
	disease diagnostics, 92
_	distribution estimation, 157
L	hedging
log-odds, 89	CDO, 79, 115, 133
logit, 89	mortgage pipeline, 87, 115, 142
logit transformation, 89	portfolio, 120, 137
Lorenz curve, 23	index tracking, 76, 78, 83
lottery, 15	portfolio inverse, 113
	portfolio replication, 85, 159
M	portfolio selection, 188, 220, 225, 233
M	Markowitz, 190, 191
margin	Markowitz-type, 106, 109, 188, 192
hard, 93	sparse reconstruction, 76, 179
separation, 93 soft, 93	sparse signal reconstruction, 80, 174 test selection, 121, 169
market portfolio, 105, 109	therapy treatment planning, 76, 184
master fund	PSG function
negative type, 108, 110	avg_cvar_risk_ni, 196
positive type, 105, 110	avg_2val_nsk_m, 190 avg, 149, 150
threshold type, 108, 110	avg_cvar_risk_ni, 200
mean-variance approach, 10, 20	avg_pm_pen_ni, 196, 203
median, 30	avg_pr_pen_ni, 196, 201
moment	avg_var_risk_ni, 196, 199
n th -order, 5	cardn, 135, 175, 176
central, 5	cdar_dev, 124, 222, 224, 233, 236, 238
partial lower, 8	cdarmulti_dev, 124, 227, 231, 233, 235, 236
partial upper, 8	cvar_comp_pos, 169, 171
2nd-order	cvar_dev, 144, 189, 192
partial lower, 184	cvar_max_risk, 121, 153
partial upper, 184	cvar_risk, 117, 140, 144, 154, 161, 190,
	195, 198, 206, 222, 227, 228, 235,
	236
N	drawdown_dev_avg, 222, 223
norm	drawdown_dev_max, 124, 222
\mathcal{L}^p , 6	drawdownmulti_dev_avg, 227, 230
vector, 135	drawdownmulti_dev_max, 124, 227, 228

PSG function (cont.)	risk identifier, 24, 25
entropyr, 157, 213, 215	risk measure
logexp_sum, 91, 217–219	\mathcal{L}^p -type, 27, 28
max_cvar_risk, 206	averse, 26, 101, 102
max_dev, 150	coherent, 27
max_var_risk, 206	coherent, 21, 26, 27
meanabs_dev, 135, 144, 145	CVaR-type, 27
meanabs_pen, 78, 135, 161, 174, 178, 181,	monotone, 26
182	semi- \mathcal{L}^p type, 27
meansquare, 181, 183, 213, 214	semi- \mathcal{L}^p type, 28
pm2_pen, 184, 185	strict expectation bounded, 26
	risk profile, 23, 28
pm_pen, 82, 84, 163–165, 167	118K profile, 23, 26
pm_pen_g, 84, 163–165, 167	
pm_pen_ni, 196	
polynom_abs, 135, 175, 178, 181, 182,	S
217–219	sample mean, 46
pr_pen_ni, 196	sample variance, 47
prmulti_pen_ni_g, 170, 172	semideviation
st_dev, 143, 146, 189–191	standard lower, 10, 21, 24, 37, 39, 40
st_pen, 76	standard lower, 10, 21, 24, 37, 39, 40 standard upper, 8, 21
var_risk, 117, 144, 195, 197, 206	11
var_risk_g, 144	skewness, 5
variance, 214, 215	SPARCO, 76, 80, 179
	statistic
	associated with error, 29
0	stochastic dominance, 9
quantile interval, 12	first-order, 9, 13
quantific interval, 12	second-order, 9, 13
	subdifferential, 24, 102, 103
D	subgradient, 24
R .	support vector, 93
regression	supremum
least median of squares (LMS), 86	essential, 3
least squares, 72–74, 162, 163	SVM, 92, 204
linear, 72	C-, 204
least-trimmed squares (LTS), 86	ν-, 94, 204, 206–209
linear, 71	extended, 95, 204
generalized, 73	robust, 95, 204, 210, 211
logistic, 89, 217	
regularized, 217	Ev-, 95, 204
median, 73, 77	hard margin, 93
mixed quantile, 83, 165, 167	linear, 92
nonlinear, 71	nonlinear, 92
nonparametric, 71	soft margin, 93
parametric, 71	
quantile, 29, 73, 81, 162, 166	
risk acceptable, 85	T
robust, 86	theorem
unbiased linear, 84	Boltzmann's, 54, 57, 69
•	
risk, 19	one-fund, 105, 109
credit, 117	tranche
default, 117	equity, 133
mortgage pipeline, 115, 142	mezzanine, 133
risk aversion, 26	senior, 133
risk envelope, 23, 25, 102	super senior, 133