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 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 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 — Michael Zabarankin
Gainesville, FL, USA — Stan Uryasev

Contents

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

$$\mathbb{P}[|X| \geq a] \leq \frac{E[|X|]}{a}, \qquad a > 0. \tag{3.1.1}$$

Its proof is straightforward:

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

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,

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

M. Zabarankin and S. Uryasev, *Statistical Decision Problems: Selected Concepts and Portfolio Safeguard Case Studies*, Springer Optimization and Its Applications 85,
DOI 10.1007/978-1-4614-8471-4_3,

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

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

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

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

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

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

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:

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

or not to drop below the threshold τ:

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

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

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

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 $\mathbb{P}[X \le \tau] \equiv F_X(\tau)$, 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 $X_1,\dots,X_n$ are a sequence of independent random variables such that $E[X_k] = 0$ and $\sigma(X_k) < \infty$, $k = 1,\dots,n$, and let $S_k = \sum_{j=1}^{k} X_j$. Then *Kolmogorov's inequality* estimates the probability of $\max\limits_{1\le k\le n} |S_k|$ to exceed a threshold a in terms of $\sigma(S_n)$:

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

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 $\mathrm{CVaR}_\alpha^\Delta$, is formulated as follows: for

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

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

$$\begin{aligned} u_{\mathscr{D}}(\delta) = &\inf_{X \in \mathscr{L}^p(\Omega)} \mathscr{D}(X) \\ &\text{subject to} \quad X \in \mathscr{U} = \{X \mid E[X] = 0,\ \mathbb{P}[X \le -a] \ge \delta\}, \end{aligned} \tag{3.2.2}$$

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 $\mathscr{D}$ over a subset of $\mathscr{U}$, 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 $\mathrm{CVaR}_\alpha^\Delta$, the function $u_{\mathscr{D}}(\delta)$ and its inverse, which is $g_{\mathscr{D}}(d)$, are given by

$$\begin{aligned} u_\sigma(\delta) &= \sigma(X^*(\delta)) = a\sqrt{\frac{\delta}{1-\delta}}, & g_\sigma(d) &= u_\sigma^{-1}(d) = \frac{d^2}{a^2 + d^2}, \\ u_{\mathrm{MAD}}(\delta) &= \mathrm{MAD}(X^*(\delta)) = 2a\delta, & g_{\mathrm{MAD}}(d) &= u_{\mathrm{MAD}}^{-1}(d) = \frac{d}{2a}, \\ u_{\sigma_-}(\delta) &= \sigma_-(X^*(\delta)) = a\sqrt{\delta}, & g_{\sigma_-}(d) &= u_{\sigma_-}^{-1}(d) = \frac{d^2}{a^2}, \end{aligned}$$

[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.

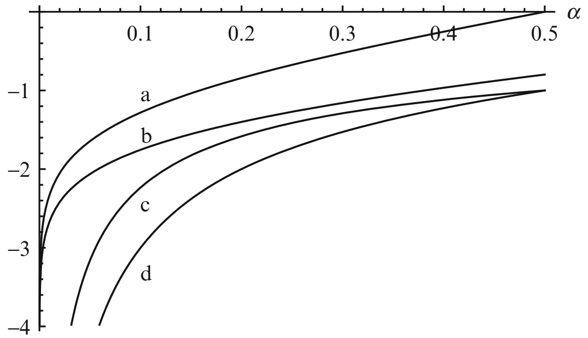

Fig. 3.1 Comparison of $q_Z^+(\alpha)$ (curve a) with $\overline{q}_Z(\alpha)$ (curve b), $-\sqrt{\alpha^{-1}-1}$ (curve c), and $-1/\sqrt{2\alpha}$ (curve d) for a standard normal random variable Z for $\alpha \in (0, 1/2]$

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

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

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

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

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 $\mathrm{CVaR}_\alpha^\Delta$ is given by

$$\mathbb{P}[|X-\mu| \geq a] \leq \frac{\mathrm{MAD}(X)}{a}, \tag{3.2.13}$$

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 $X \in \mathscr{L}^m(\Omega)$ provided that the first m moments of X are known to be $\mu_1 \in \mathbb{R}, \ldots, \mu_m \in \mathbb{R}$:

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

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

M. Zabarankin and S. Uryasev, *Statistical Decision Problems: Selected Concepts and Portfolio Safeguard Case Studies*, Springer Optimization and Its Applications 85,
DOI 10.1007/978-1-4614-8471-4_5,

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:

$$\begin{aligned} p_1 = \ldots = p_n &= \frac{1}{n} \quad \text{in discrete case,} \\ f_X(t) &= \frac{1}{b-a} I_{\{t \in [a,b]\}} \quad \text{in continuous case.} \end{aligned} \tag{5.1.6}$$

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

$$p_k = \frac{e^{\rho x_k}}{\sum_{k=1}^n e^{\rho x_k}}, \quad k = 1, \ldots, n, \tag{5.1.7}$$

and

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

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 $k = 1, 2, \ldots$ and $\mu > 1$, then $e^\rho = 1 - 1/\mu$ and the maximum entropy distribution (5.1.7) reduces to

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

(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

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

where ρ_1 and ρ_2 are found from the system

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

whereas (5.1.5) simplifies to

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

with λ_1 and λ_2 found from the system

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

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

$$\begin{aligned} &\max_{f_X(t)} \; - \int_V f_X(t) \ln f_X(t)\, dt \\ &\text{subject to} \int_V h_k(t)\, f_X(t)\, dt = a_k, \quad k = 1, \ldots, m, \\ &\qquad\qquad \int_V f_X(t)\, dt = 1, \qquad f_X(t) \geq 0, \quad t \in V, \end{aligned} \tag{5.1.11}$$

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

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

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

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 $p_1, \dots, p_n$. 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: $\sum_{i=1}^{n} a_{ij}\, p_i = 0$, $j = 1, \dots, m$. 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 $\sum_{i=1}^{n} \underline{a}_{ij}\, p_i \le 0$ and $\sum_{i=1}^{n} \overline{a}_{ij}\, p_i \ge 0$, $j = 1, \dots, m$, and the problem is to maximize the Shannon entropy with respect to the hazard rate scenario probabilities $p_1, \dots, p_n$ subject to the no-arbitrage constraints:

$$\begin{aligned}
&\max_{p_1,\dots,p_n} \; -\sum_{i=1}^{n} p_i \ln p_i \\
&\text{subject to} \quad \sum_{i=1}^{n} p_i = 1, \quad p_i \ge 0, \quad i = 1, \dots, n, \\
&\qquad \left.\begin{aligned}
&\sum_{i=1}^{n} \underline{a}_{ij}\, p_i \le 0, \quad j = 1, \dots, m, \\
&\sum_{i=1}^{n} \overline{a}_{ij}\, p_i \ge 0, \quad j = 1, \dots, m.
\end{aligned}\right\} \quad \text{no-arbitrage constraints}
\end{aligned} \tag{5.1.13}$$

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):

$$
\begin{aligned}
&\max_{p_1,\dots,p_n} \quad -\sum_{i=1}^{n} p_i \ln p_i \\
&\text{subject to} \quad \sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0, \quad i = 1,\dots,n, \\
&\qquad \left.\begin{aligned}
&\sum_{i=1}^{n} \underline{a}_{ij}\, p_i \leq 0, \quad j = 1,\dots,m, \\
&\sum_{i=1}^{n} \overline{a}_{ij}\, p_i \geq 0, \quad j = 1,\dots,m,
\end{aligned}\right\} \quad \text{no-arbitrage constraints} \\
&\qquad 1 \leq w_l \leq w_r \leq n, \\
&\qquad \frac{p_{i-1} + p_{i+1}}{2} \geq p_i, \quad i = 2,\dots,w_l - 1 \qquad \text{(left slope is convex)} \\
&\qquad \frac{p_{i-1} + p_{i+1}}{2} \leq p_i, \quad i = w_l + 1,\dots,w_r - 1 \quad \text{(hump is concave)} \\
&\qquad \frac{p_{i-1} + p_{i+1}}{2} \geq p_i, \quad i = w_r + 1,\dots,n - 1 \quad \text{(right slope is convex)}
\end{aligned}
\tag{5.1.14}
$$

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*:

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

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

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, $D_{\mathrm{KL}}(Y||X) \neq D_{\mathrm{KL}}(X||Y)$, 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 $D_{\mathrm{KL}}(X||Y)$ for a given reference random variable Y subject to any additional constraints on X:

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

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

$$\begin{aligned} &\min_{p_1,\ldots,p_n} \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \\ &\text{subject to } \sum_{i=1}^{n} p_i = 1, \quad Ap \leq b, \\ &\qquad l_i \leq p_i \leq u_i, \quad i = 1, \ldots, n, \end{aligned} \tag{5.2.1}$$

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 $0 \leq l_i \leq u_i \leq 1$, $i = 1, \ldots, n$.

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 $X \in \mathscr{L}^1(\Omega)$, whose mean and law-invariant deviation $\mathscr{D} : \mathscr{L}^p(\Omega) \mapsto [0, \infty]$, $p \in [1, \infty]$, are known to be μ and d, respectively, is formulated by

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

Let $X_0 \in \mathscr{L}^1(\Omega)$ 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

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

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

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 $\mathrm{CVaR}_\alpha^\Delta$ 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 $Y \in \mathscr{L}^1(\Omega)$ subject to a constraint on the deviation $\mathscr{D}$ projected from a nondegenerate error measure $\mathscr{E}$ and subject to a constraint on the statistic $\mathscr{S}$ associated with $\mathscr{E}$:

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

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

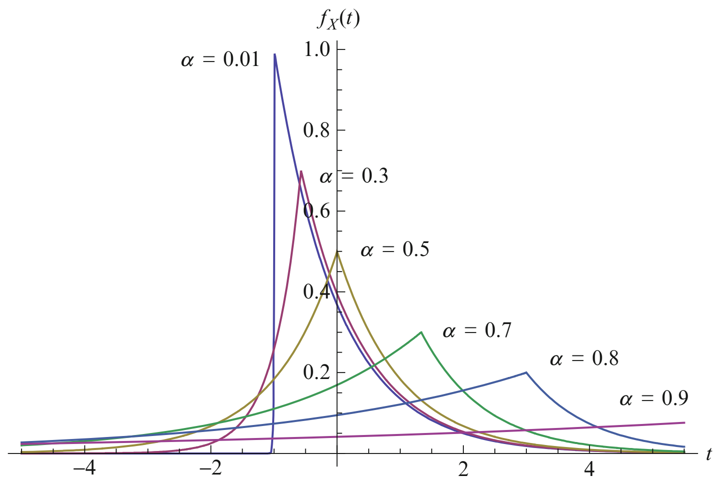

Fig. 5.1 The PDF $f_X(t)$ that maximizes the Shannon entropy $S(X)$ subject to constraints on the mean and CVaR deviation: $E[X]=0$ and $\mathrm{CVaR}_\alpha^\Delta(X)=1$ for $\alpha=0.01$, 0.3, 0.5, 0.7, 0.8, and 0.9

Let $Y_0 \in \mathscr{L}^1(\Omega)$ 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

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

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

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

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

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

where $\mathscr{S}$ is the statistic associated with $\mathscr{E}$ 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 $C_0 \in \mathscr{S}(X_0)$. Then (5.4.2) reduces to (5.4.4) by substitution $Y_0 = X_0 - C_0$. Indeed, $S(X_0) = S(Y_0)$, $0 \in \mathscr{S}(Y_0)$, and $\mathscr{D}(X_0) = \mathscr{D}(Y_0)$.

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 $C^* \in \mathscr{S}(Z^*)$, and let $Y_0^* = Z^* - C^*$, then $\mathscr{E}(Y_0^*) = 1$. Indeed, by definition of the statistic $\mathscr{S}$ associated with $\mathscr{E}$,

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

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 $0 \in \mathscr{S}(Y_0)$, 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, $0 \in \mathscr{S}(Y_0^*)$ and $\mathscr{E}(Y_0^*) = 1$, 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

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

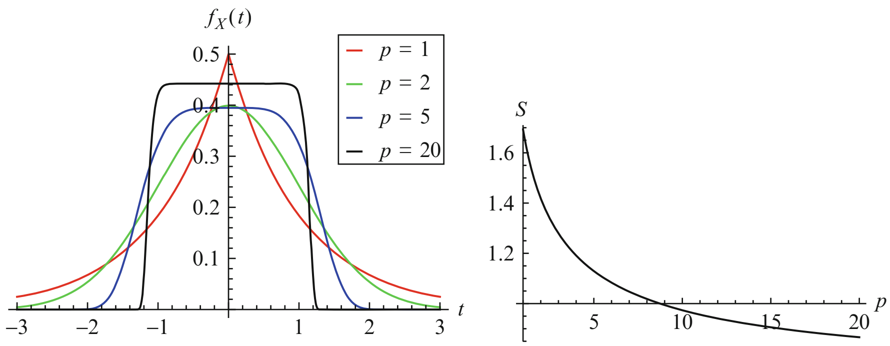

Fig. 5.2 PDFs $f_Z(t)$ that maximize the Shannon entropy $S(Z)$ subject to the constraint $\|Z\|_p = 1$ for $p = 1, 2, 5, 20$ and the value of the entropy $S(Z)$ for these PDFs as a function of $p \in [1, \infty)$. The PDF with the sharp spike corresponds to $p = 1$, and the PDF dispersion decreases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

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

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, $\lim_{p\to\infty} f_{Z^*}(t) = \frac{1}{2} I_{\{-1 \le t \le 1\}}$ 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 $\mathscr{D} = \text{MAD}$ 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 $\mathscr{E}(Z) = \mathscr{E}_\alpha(Z)$ 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

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

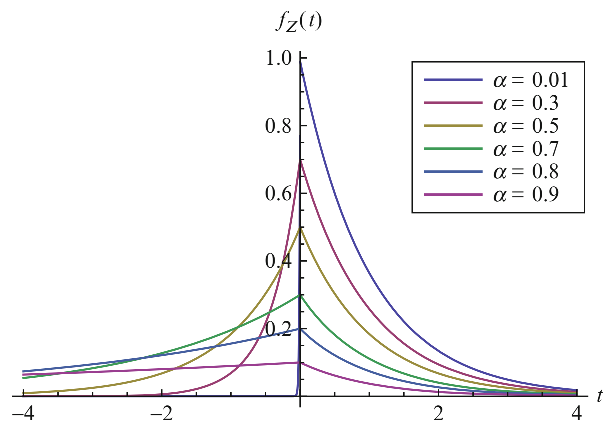

Fig. 5.3 The PDF $f_Z(t)$ that maximizes the Shannon entropy $S(Z)$ subject to a constraint on the asymmetric mean absolute error (2.0.3): $\mathscr{E}_\alpha(Z) = 1$ for $\alpha = 0.01, 0.3, 0.5, 0.7, 0.8$, and 0.9

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 $\mathscr{E}_\alpha$ and the statistic $\mathscr{S}$ associated with $\mathscr{E}_\alpha$ are CVaR deviation and α-quantile, respectively. Let X_0^* solve (5.4.2) with $\mathscr{D} = \mathrm{CVaR}_\alpha^\Delta$:

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

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

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

(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 (2006)
[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)

M. Zabarankin and S. Uryasev, *Statistical Decision Problems: Selected Concepts and Portfolio Safeguard Case Studies*, Springer Optimization and Its Applications 85, DOI 10.1007/978-1-4614-8471-4,

[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 mean-semideviation 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)

[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)

Index

M. Zabarankin and S. Uryasev, *Statistical Decision Problems: Selected Concepts and Portfolio Safeguard Case Studies*, Springer Optimization and Its Applications 85,
DOI 10.1007/978-1-4614-8471-4,