HUMBOLDT-UNIVERSITÄT ZU BERLIN
Institut für Mathematik

Humboldt Distinguished Lecture Series in Applied Mathematics

Risk and Uncertainty in Optimization

R. Tyrrell Rockafellar

This lecture series is intended for graduate students in mathematics and economics with an interest in optimization and finance. It is given by a pioneer in optimization and convex analysis and takes place:

January 8th; 11:00 - 13:00 and 15:00 - 17:00;
Erwin Schrödinger Zentrum; Room 0.307.

January 9th, 11:00 - 13:00 and 15:00 - 17:00;
Johann v. Neumann Haus; Room 1.013.

The lectures cover an array of topics from convex analysis, optimization and risk theory:

- Optimization Modeling with Convexity and Duality
- Risk Measures and Safeguarding in Optimization
- Deviation Measures and Generalized Linear Regression
- Utility, Generalized Entropy and Measures of Liability

There is no registration. For further information and course material, please visit

www.math.hu-berlin.de/~horst/

Organizer:

Ulrich Horst
Deutsche Bank Professor of Applied Financial Mathematics

Humboldt-Universität zu Berlin
Institut für Mathematik
Unter den Linden 6
D-10099 Berlin

In collaboration with the *Berlin Mathematical School* and the *Quantitative Products Laboratory, a joint venture of Deutsche Bank AG, Technische Universität Berlin and Humboldt Universität zu Berlin.*

email: horst@math.hu-berlin.de

R. Tyrrell Rockafellar is Professor Emeritus at the University of Washington where he pioneered in the mathematics of optimization and its many applications. He is currently also associated with the University of Florida for collaborations in the theory of risk. His awards include the *Dantzig Prize* (1982), *The Lanchester Prize* (1998), the *von Neumann Theory Prize* (1999), and honorary doctorates from several universities. Among more than 200 publications are his books "Convex Analysis" (1970) and "Variational Analysis" (1998) which have long become standard references.

OPTIMIZATION MODELING WITH CONVEXITY AND DUALITY

Terry Rockafellar
University of Washington, Seattle
University of Florida, Gainesville

Humboldt University, Berlin —— January, 2009

LECTURE 1

Basic Framework of Optimization

problems of "continuous" rather than "discrete" type

$\mathcal{X}$ some linear space, e.g., $\boldsymbol{R}^n$ or $\mathcal{L}^p$(probability space)

$f : \mathcal{X} \to \bar{\boldsymbol{R}} = [-\infty, \infty]$ some function

$$\operatorname{dom} f = \{ x \in \mathcal{X} \mid f(x) < \infty \} \quad \text{effective domain}$$
$$\operatorname{epi} f = \{ (x, \alpha) \in \mathcal{X} \times \boldsymbol{R} \mid f(x) \le \alpha \} \quad \text{epigraph}$$

Abstract model in optimization

$(\mathcal{P})$ minimize $f(x)$ over all $x \in \mathcal{X}$

feasible solutions: $x \in \operatorname{dom} f$

optimal solutions: $x \in \operatorname{argmin} f$ $\quad \operatorname{argmin}(\mathcal{P})$

optimal value: $\inf f$ $\quad \inf(\mathcal{P})$

convex case: f convex, meaning that $\operatorname{epi} f$ is a convex set

$$f((1-\tau)x' + \tau x'') \le (1-\tau)f(x') + \tau f(x'') \text{ for } \tau \in (0,1)$$

Parametric Embedding and Sensitivity

$\mathcal{U}$ = some linear space of perturbations u

$F : \mathcal{X} \times \mathcal{U} \to \bar{R}$ some function with $F(x, 0) = f(x)$

Parameterized model in optimization

$(\mathcal{P}(u))$ minimize $F(x, u)$ over all $x \in \mathcal{X}$

$(\mathcal{P}(0)) = (\mathcal{P})$

convex parameterization: $F(x, u)$ convex in u

full convexity: $F(x, u)$ convex jointly in x and u

Optimal value function

$p(u) = \inf(\mathcal{P}(u)) = \inf_x F(x, u)$, with $p(0) = \inf(\mathcal{P})$

full convexity $\Longrightarrow$ p is convex

sensitivity to perturbations: generalized derivatives of p at 0

Example of Nonlinear Programming

problem model:

minimize $c_0(x)$ over $x \in S$ having $c_i(x) \le 0$ for $i = 1, \ldots, m$

$S \subset \mathcal{X}, \quad c_i : S \to R$ for $i = 0, 1, \ldots, m$

corresponding objective in abstract format:

$f(x) = c_0(x)$ if $x \in S$ and $c_i(x) \le 0$ for $i = 1, \ldots, m$

but otherwise $f(x) = \infty$

canonical parameterization: $u = (u_1, \ldots, u_m)$

$F(x,u) = c_0(x)$ if $x \in S$ and $c_i(x) + u_i \le 0$ for $i = 1, \ldots, m$

but otherwise $F(x,u) = \infty$

Observations:

- f is convex if S = convex set and each c_i = convex function
- $F(x,u)$ is always convex in u
- $F(x,u)$ is jointly convex in x and u when f is convex.

Example of Composite Objectives

problem model: minimize $\theta(g_1(x), \ldots, g_d(x))$ over all $x \in S$

$S \subset \mathcal{X}$, $c_i : \mathcal{X} \to \boldsymbol{R}$, $\theta : \boldsymbol{R}^d \to (-\infty, \infty]$ convex nondecreasing

corresponding objective function in abstract format:

$$f(x) = \theta(g_1(x), \ldots, g_d(x)) \text{ if } x \in S$$
$$\text{but otherwise } f(x) = \infty$$

canonical parameterization: $u = (u_1, \ldots, u_d)$

$$F(x,u) = \theta(g_1(x) + u_1, \ldots, g_d(x) + u_d) \text{ if } x \in SX$$
$$\text{but otherwise } F(x,u) = \infty$$

Observations:

- f is convex when $S =$ convex set, each $g_i =$ convex function
- $F(x,u)$ is always convex in u
- $F(x,u)$ is jointly convex in x and u when f is convex.

Example of Stochastic Programming

$(\Omega, \mathcal{F}, P)$ = probability space of future states ω

One-stage model

minimize $\Phi(x_0) = E_\omega\{f(x_0, \omega)\}$ over all $x_0 \in \mathcal{X}_0$

$f : \mathcal{X}_0 \times \Omega \to \bar{R}$ incorporates constraints!

$\Phi(x_0) < \infty$ will require $f(x_0, \omega) < \infty$ a.s. in ω

(various technicalities involving measurability need attention)

Two-stage model

minimize $\Phi(x_0, x_1(\cdot)) = E_\omega\{f(x_0, x_1(\omega), \omega)\}$ over all

$x_0 \in \mathcal{X}_0$ and [measurable] mappings $x_1(\cdot) : \Omega \to \mathcal{X}_1$

$x_1(\omega)$ = recourse decision

The expectation functionals Φ are special **integral functionals**

Φ inherits convexity from the integrand f

Lagrangians and Dual Problems

primal problem $(\mathcal{P})$: minimize $f(x)$ over $x \in \mathcal{X}$

Lagrangian for $(\mathcal{P})$ and a multiplier space $\mathcal{Y}$

any function L on $\mathcal{X} \times \mathcal{Y}$ having

$$f(x) = \sup_{y \in \mathcal{Y}} L(x,y) \text{ for all } x \in \mathcal{X}$$

let $g(y) = \inf_{x \in \mathcal{X}} L(x,y)$ for all $y \in \mathcal{Y}$

dual problem $(\mathcal{D})$: maximize $g(y)$ over all $y \in \mathcal{Y}$,

Basic primal-dual relationships

(a) $\inf(\mathcal{P}) \geq \sup(\mathcal{D})$ always

(b) $\left[\inf(\mathcal{P}) = \sup(\mathcal{D}),\ \bar{x} \in \operatorname{argmin}(\mathcal{P}),\ \bar{y} \in \operatorname{argmax}(\mathcal{D}) \right]$

$\iff \left[\inf_x L(x,\bar{y}) = L(\bar{x},\bar{y}) = \sup_y L(\bar{x},y) \right]$ saddle point

saddle point existence: unlikely unless $L(x,y)$ is convex-concave

Paired Spaces for Developing Duality

linear spaces $\mathcal{U}$ and $\mathcal{Y}$, with bilinear form $\langle u, y\rangle$ on $\mathcal{U} \times \mathcal{Y}$

Compatible topologies

the continuous linear functionals on $\mathcal{U}$ are $u \longrightarrow \langle u, y\rangle$ for $y \in \mathcal{Y}$
the continuous linear functionals on $\mathcal{Y}$ are $y \longrightarrow \langle u, y\rangle$ for $u \in \mathcal{U}$

Examples:

- $\mathcal{U} = \boldsymbol{R}^m,\ \mathcal{Y} = \boldsymbol{R}^m,\ \langle u, y\rangle = u{\cdot}y = \sum_{i=1}^m u_i y_i$ usual topology
- $\mathcal{U} = \mathcal{L}^p_m(\Omega, \mathcal{F}, P),\ \mathcal{Y} = \mathcal{L}^q_m(\Omega, \mathcal{F}, P)$, usual pairing, with $\langle u, y\rangle = E\{u{\cdot}y\} = \int_\Omega \sum_{i=1}^m u_i(\omega) y_i(\omega) dP(\omega)$
  the norm topologies, except for $\mathcal{L}^\infty$ the weak* topology
- the weak topologies $\sigma(\mathcal{U}, \mathcal{Y})$ on $\mathcal{U}$ and $\sigma(\mathcal{Y}, \mathcal{U})$ on $\mathcal{Y}$

Note: the **closed convex** sets and **lsc convex** functions (lower semicontinuous) are **the same in all compatible topologies**

Conjugate Convex Functions

$\mathcal{U}$ and $\mathcal{Y}$: paired linear spaces with compatible topologies

Legendre-Fenchel transform

$\varphi : \mathcal{U} \to \bar{R}$ any function
$\varphi^* : \mathcal{Y} \to \bar{R}$ its conjugate, $\varphi^*(y) = \sup_u \{\langle u, y\rangle - \varphi(u)\}$
$\varphi^{**} : \mathcal{U} \to \bar{R}$ its biconjugate, $\varphi^{**}(u) = \sup_y \{\langle u, y\rangle - \varphi^*(y)\}$

Closed* convex functions (lsc and $> -\infty$, unless $\equiv -\infty$)

- φ^* is a closed* convex function
- φ^{**} is the largest closed* convex function $\leq \varphi$

Conjugacy correspondence

The **closed* convex** functions φ on $\mathcal{U}$ and ψ on $\mathcal{Y}$ correspond **one-to-one** to each other under: $\psi = \varphi^*$, $\varphi = \psi^*$

The constant functions ∞ and $-\infty$ are conjugate to each other

Conjugate Duality Scheme in Optimization

$\mathcal{U}$ and $\mathcal{Y}$: paired linear spaces with compatible topologies

For the problem $(\mathcal{P})$ of minimizing $f(x)$ over $x \in \mathcal{X}$, consider

- parameterizations $F : \mathcal{X} \times \mathcal{U} \to \bar{R}$ with $F(x, \cdot)$ closed* convex
- Lagrangians $L : \mathcal{X} \times \mathcal{Y} \to \bar{R}$ with $-L(x, \cdot)$ closed* convex

Parameterizations versus Lagrangians

Such F and L correspond to each other **one-to-one** under

$$L(x, y) = \inf_u \{F(x, u) - \langle u, y \rangle\}, \quad F(x, u) = \sup_u \{L(x, y) + \langle u, y \rangle\}$$

$$F(x, u) \text{ convex in } (x, u) \iff L(x, y) \text{ concave in } y$$

Nonlinear programming example: $u \in R^m$, $y \in R^m$

$$F(x, u) = c_0(x) \text{ if } x \in S \text{ and } c_i(x) + u_i \leq 0 \text{ for } i = 1, \dots, m$$
$$\text{but otherwise } F(x, u) = \infty$$

$$L(x, y) = c_0(x) + y_1 c_1(x) + \cdots + y_m c_m(x) \text{ if } x \in S,\ y \geq 0$$
$$\text{and } = \infty \text{ if } x \notin S,\ y \geq 0, \text{ but } = -\infty \text{ if } y \not\geq 0$$

Main Results for the Conjugate Duality Scheme

$\mathcal{U}$ and $\mathcal{Y}$: paired linear spaces with compatible topologies

$$\text{Lagrangian } L(x,y) \quad \leftrightarrow \quad \text{parameterization } F(x,u)$$

$(\mathcal{P})$ minimize $f(x)$ over $x \in X$ where $f(x) = \sup_y L(x,y)$

$(\mathcal{D})$ maximize $g(y)$ over $y \in Y$ where $g(y) = \inf_x L(x,y)$

Optimal value function:

$$p(u) = \inf_x F(x,u) = \inf(\mathcal{P}(u)) \qquad \text{where } F(x,0) = f(x)$$

Characterization of primal-dual optimal values and solutions

(a) $\inf(\mathcal{P}) = p(0), \quad \sup(\mathcal{D}) = p^{**}(0)$

(b) $(\bar{x},\bar{y})$ is a saddle point of $L(x,y)$ if and only if
$\bar{x} \in \mathrm{argmin}(\mathcal{P})$ and $p(u) \geq p(0) + \langle u, \bar{y}\rangle$ for all $u \in \mathcal{U}$

Key question: when does there exist $\bar{y}$ with this relation to p at 0?

Subgradients and Directional Derivatives

Subgradients of convex analysis

For $\varphi : \mathcal{U} \to \bar{R}$, $\varphi \not\equiv \infty$, $u \in \mathcal{U}$, $y \in \mathcal{Y}$:

$$y \in \partial\varphi(u) \;\text{ means }\; \varphi(u+w) \geq \varphi(u) + \langle w, y\rangle \text{ for all } w \in \mathcal{U}$$

Directional derivatives of convex functions

For φ convex on $\mathcal{U}$, finite at $\bar{u}$, bounded above around $\bar{u}$:

(a) $\varphi'(\bar{u}; w) = \lim_{\tau \to 0^+} \dfrac{\varphi(\bar{u}+\tau w) - \varphi(\bar{u})}{\tau}$ is finite, convex in w

(b) $\varphi'(\bar{u}; w) = \max\{\langle w, y\rangle \mid y \in \partial\varphi(\bar{u})\}$

(c) for R^n: $\partial\varphi(\bar{u}) = \{\bar{y}\} \iff \varphi$ diff. at $\bar{u}$ with $\bar{y} = \nabla\varphi(\bar{u})$

Relation to conjugacy

For conjugate functions φ on $\mathcal{U}$ and ψ on $\mathcal{V}$, not $\equiv \infty$ or $\equiv -\infty$:

(a) $\varphi(u) + \varphi(y) \geq \langle u, y\rangle$ for all $u \in \mathcal{U}$ and $y \in \mathcal{Y}$

(b) equality holds for $u, y \iff y \in \partial\varphi(u) \iff u \in \partial\psi(y)$

Fenchel-Type Duality Schemes

$\mathcal{U} \leftrightarrow \mathcal{Y}$, $\mathcal{X} \leftrightarrow \mathcal{V}$: paired linear spaces with compatible topologies

proper lsc convex h on $\mathcal{X}$, k on $\mathcal{U}$, conjugates h^* on $\mathcal{V}$, k^* on $\mathcal{Y}$

$c \in \mathcal{V}$, $b \in \mathcal{U}$, continuous linear $A : \mathcal{X} \to \mathcal{U}$, adjoint $A^* : \mathcal{Y} \to \mathcal{V}$

primal $(\mathcal{P})$ $\min f(x) = \langle c, x\rangle + h(x) + k(b - Ax)$ over $x \in \mathcal{X}$

dual $(\mathcal{D})$ $\max g(y) = \langle b, y\rangle - k^*(y) - h^*(A^*y - c)$ over $y \in \mathcal{Y}$

Lagrangian: $L(x, y) = \langle c, x\rangle + h(x) + \langle b, y\rangle - k^*(y) - \langle Ax, y\rangle$

feasibility in $(\mathcal{P})$ $\iff$ $b \in [A\,\mathrm{dom}\,h + \mathrm{dom}\,k]$

feasibllity in $(\mathcal{D})$ $\iff$ $c \in [A^*\,\mathrm{dom}\,k^* - \mathrm{dom}\,h^*]$

Duality Theorem

Suppose $\mathcal{U}$ and $\mathcal{V}$ are Banach (in the compatible topologies!)

(a) $\inf(\mathcal{P}) = \max(\mathcal{D}) < \infty$ if $b \in \mathrm{int}[A\,\mathrm{dom}\,h + \mathrm{dom}\,k]$

(b) $\min(\mathcal{P}) = \sup(\mathcal{D}) > -\infty$ if $c \in \mathrm{int}[A^*\,\mathrm{dom}\,k^* - \mathrm{dom}\,h^*]$

Some General References

[1] R. T. Rockafellar (1974), *Conjugate Duality and Optimization*, No. 16 in the Conference Board of Math. Sciences Series, SIAM Publications, Philadelphia (74 pages)

[2] R. T. Rockafellar (1970), *Convex Analysis*, Princeton University Press, Princeton, New Jersey (available from 1997 also in paperback in the series Princeton Landmarks in Mathematics).

[3] R. T. Rockafellar, R. J.-B. Wets (1998, 2005), *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften 317, Springer-Verlag, Berlin (second printing, with corrections: 2005)

[4] R. T. Rockafellar (1999), "Extended nonlinear programming," in *Nonlinear Optimization and Related Topics* (G. Di Pillo and F. Giannessi, eds,), Kluwer, 381-399 downloadable

[5] R. T. Rockafellar (1993), "Lagrange multipliers and optimality," *SIAM Review* 35, 183–238

DOWNLOADS

website: www.math.washington.edu/∼rtr/mypage.html

Available besides [4] and some other relatively recent papers:

- Course lecture notes on *Fundamentals of Optimization*

Very introductory material in finite dimensions, which nonetheless covers geometric nonsmooth analysis and optimality conditions in terms of normal cones, as well as properties of polyhedrality

- Course lecture notes on *Optimization Under Uncertainty*,

The basics of traditional stochastic programming, without use of "risk measures," but with duality and a build-up to multistage models in a framework of scenarious and decomposition

RISK MEASURES AND SAFEGUARDING IN OPTIMIZATION

Terry Rockafellar
University of Washington, Seattle
University of Florida, Gainesville

Humboldt University, Berlin —— January, 2009

LECTURE 2

Uncertainty in Optimization

Decisions (optimal?) must be taken before the facts are all in

- A bridge must be built to withstand floods, wind storms or earthquakes
- A portfolio must be purchased with incomplete knowledge of how it will perform
- A product's design constraints must be viewed in terms of "safety margins"

What are the consequences for optimization?
How may this affect the way problems are formulated and solved?

How can "risk" properly be taken into account, with attention paid to the attitudes of the optimizer?

How should the future, where the essential uncertainty resides, be modeled with respect to decisions and information?

The Fundamental Difficulty Caused by Uncertainty

with simple modeling of the future

A standard form of optimization problem without uncertainty:

minimize $c_0(x)$ over all $x \in S$ satisfying $c_i(x) \leq 0, \; i = 1, \ldots, m$
for a set $S \subset \boldsymbol{R}^n$ and functions $c_i : S \mapsto \boldsymbol{R}$

Incorporation of future states $\omega \in \Omega$ in the model:
the decision x must be taken before ω is known

Choosing $x \in S$ no longer fixes numerical values $c_i(x)$, but only fixes **functions on** Ω: $\underline{c}_i(x) : \omega \mapsto c_i(x, \omega), \quad i = 0, 1, \ldots, m$

Optimization objectives and constraints must be reconstrued in terms of such function, but how? There is no universal answer...

Various approaches: old/new? good/bad? yet to be discovered?
Adaptations to attitudes about "risk"?

Example: Linear Programming Context

Problem without uncertainty: $c_i(x) = a_{i1}x_1 + \cdots a_{in}x_n - b_i$

minimize $a_{01}x_1 + \cdots a_{0n}x_n - b_0$ over $x = (x_1, \ldots, x_n) \in S$

subject to $a_{i1}x_1 + \cdots a_{in}x_n - b_i \leq 0$ for $i = 1, \ldots, m$,

where $S = \{x \mid x_1 \geq 0, \ldots, x_n \geq 0$ & other conditions? $\}$

Effect of uncertainty: $c_i(x, \omega) = a_{i1}(\omega)x_1 + \cdots a_{in}(\omega)x_n - b_i(\omega)$

Portfolio illustration with financial instruments $j = 1, \ldots, n$

$r_j(\omega) =$ rate of return, $x_j =$ weight in the portfolio

portfolio rate of return $= x_1 r_1(\omega) + \cdots + x_n r_n(\omega)$

Constraints: $x \in S = \{(x_1, \ldots, x_n) \mid x_j \geq 0,\ x_1 + \cdots + x_n = 1\}$

Uncertain ingredients to incorporate in optimization model:

$c_0(x, \omega) = -[x_1 r_1(\omega) + \cdots + x_n r_n(\omega)]$

(conversion to “cost” orientation for minimization)

$c_1(x, \omega) = q(\omega) - [x_1 r_1(\omega) + \cdots + x_n r_n(\omega)]$, $q =$ benchmark

(shortfall below benchmark, desired outcome ≤ 0)

Probabilistic Framework — Random Variables

Future state space Ω modeled with a probability structure:

$$(\Omega, \mathcal{F}, P), \quad P = \text{probability measure}$$

"true"? "subjective"? or merely for reference?

Functions $X : \Omega \to R$ interpreted then as **random variables**:

cumulative distribution function $F_X : (-\infty, \infty) \to [0, 1]$

$$F_X(z) = P\{\omega \,|\, X(\omega) \leq z\}$$

expected value EX = mean value $= \mu(X)$

variance $\sigma^2(X) = E[(X - \mu(X))^2]$, standard deviation $\sigma(X)$

Technical restriction imposed here: $X \in \mathcal{L}^2$, meaning $E[X^2] < \infty$

Corresponding convergence criterion as $k = 1, 2, \ldots \infty$:

$$X_k \to X \iff \mu(X_k - X) \to 0 \text{ and } \sigma(X_k - X) \to 0$$

The functions $\underline{c}_i(x) : \omega \to c_i(x, \omega)$ are placed now in this picture:

choosing $x \in S$ yields random variables $\underline{c}_0(x), \underline{c}_1(x), \ldots, \underline{c}_m(x)$

No-Distinction Principle for Objectives and Constraints

Is there an intrinsic reason why uncertainty/risk in an objective should be treated differently than uncertainty/risk in a constraint?

NO, because of well known, elementary reformulations

Given an optimization problem in standard format:

$$\text{minimize } c_0(x) \text{ over } x \in S \text{ with } c_i(x) \leq 0, \; i = 1, \ldots, m$$

augment $x = (x_1, \ldots, x_n)$ by another variable x_{n+1}, and in terms of

$\tilde{x} = (x, x_{n+1}) \in \tilde{S} = S \times \boldsymbol{R}$,

$\tilde{c}_i(\tilde{x}) = c_i(x)$ for $i = 1, \ldots, m$,

$\tilde{c}_0(\tilde{x}) = x_{n+1}$, $\quad \tilde{c}_{m+1}(\tilde{x}) = c_0(x) - x_{n+1}$

pass equivalently to the reformulated problem:

$$\text{minimize } \tilde{c}_0(\tilde{x}) \text{ over } \tilde{x} \in \tilde{S} \text{ with } \tilde{c}_i(\tilde{x}) \leq 0, \; i = 1, \ldots, m, m+1$$

Uncertainty in $c_0, c_1, \ldots, c_m$ will not affect the objective with $\tilde{c}_0$.
It will only affect the constraints with $\tilde{c}_1, \ldots, \tilde{c}_m, \tilde{c}_{m+1}$.

Some Traditional Approaches

Aim: recapturing optimization in the face of $\underline{c}_i(x) : \omega \to c_i(x, \omega)$
each approach followed uniformly, for emphasis in illustration

Approach 1: guessing the future

- identify $\bar{\omega} \in \Omega$ as the "best estimate" of the future
- minimize over $x \in S$:

$$c_0(x, \bar{\omega}) \text{ subject to } c_i(x, \bar{\omega}) \leq 0, \; i = 1, \ldots, m$$

- pro/con: simple and attractive, but dangerous—no hedging

Approach 2: worst-case analysis, "robust" optimization

- focus on the worst that might come out of each $\underline{c}_i(x)$:
- minimize over $x \in S$:

$$\sup_{\omega \in \Omega} c_0(x, \omega) \text{ subject to } \sup_{\omega \in \Omega} c_i(x, \omega) \leq 0, \; i = 1, \ldots, m$$

- pro/con: avoids probabilities, but expensive—maybe infeasible

Approach 3: relying on means/expected values

- focus on average behavior of the random variables $\underline{c}_i(x)$
- minimize over $x \in S$:

$$\mu(\underline{c}_0(x)) = E_\omega c_0(x,\omega) \text{ subject to}$$
$$\mu(\underline{c}_i(x)) = E_\omega c_i(x,\omega) \le 0, \; i = 1,\ldots,m$$

- pro/con: common for objective, but foolish for constraints?

Approach 4: safety margins in units of standard deviation

- improve on expectations by bringing standard deviations into consideration
- minimize over $x \in S$: for some choice of coefficients $\lambda_i > 0$

$$\mu(\underline{c}_0(x)) + \lambda_0\,\sigma(\underline{c}_0(x)) \text{ subject to}$$
$$\mu(\underline{c}_i(x)) + \lambda_i\,\sigma(\underline{c}_i(x)) \le 0, \; i = 1,\ldots,m$$

- pro/con: looks attractive, but a serious flaw will emerge

The idea here: find the lowest z such that, for some $x \in S$,
$\underline{c}_0(x) - z,\ \underline{c}_1(x),\ldots,\underline{c}_m(x)$ will be ≤ 0 except in λ_i-upper tails

Approach 5: specifying probabilities of compliance

- choose probability levels $\alpha_i \in (0,1)$ for $i = 0, 1, \ldots, m$
- find lowest z such that, for some $x \in S$, one has

$$P\{\underline{c}_0(x) \leq z\} \geq \alpha_0, \quad P\{\underline{c}_i(x) \leq 0\} \geq \alpha_i \text{ for } i = 1, \ldots, m$$

- pro/con: popular and appealing, but flawed and controversial
  - no account is taken of the seriousness of violations
  - technical issues about the behavior of these expressions

Example: with $\alpha_0 = 0.5$, the median of $\underline{c}_0(x)$ would be minimized

Additional modeling ideas:

- Staircased variables: $\underline{c}_i(x)$ propagated to $\underline{c}_i^k(x) = \underline{c}_i(x) - d_i^k$

for a series of thresholds d_i^k, $k = 1, \ldots, r$ with different compliance conditions placed on having these "subvariables" $\underline{c}_i^k(x)$ be ≤ 0

- Expected penalty expressions like $E[\psi(\underline{c}_0(x))]$
- Stochastic programming, dynamic programming

Quantification of Risk

How can the "risk" be measured in a random variable X?

orientation: $X(\omega)$ stands for a "cost" or loss

negative costs correspond to gains/rewards

- Idea 1: assess the "risk" in X in terms of how **uncertain** X is:
  $\longrightarrow$ **measures $\mathcal{D}$ of deviation from constancy**
- Idea 2: capture the "risk" in X by a **numerical surrogate** for overall cost/loss: $\longrightarrow$ **measures $\mathcal{R}$ of potential loss**

$\longrightarrow$ our concentration, for now, will be on Idea 2

A General Approach to Uncertainty in Optimization

In the context of the numerical values $c_i(x) \in \boldsymbol{R}$ being replaced by random variables $\underline{c}_i(x) \in \mathcal{L}^2$ for $i = 0, 1, \ldots, m$:

- choose risk measures $\mathcal{R}_i$ of potential loss,
- define the functions $\bar{c}_i$ on $\boldsymbol{R}^n$ by $\bar{c}_i(x) = \mathcal{R}_i(\underline{c}_i(x))$, and then
- minimize $\bar{c}_0(x)$ over $x \in S$ subject to $\bar{c}_i(x) \le 0$, $i = 1, \ldots, m$.

Basic Guidelines

For a functional $\mathcal{R}$ that assigns to each random "cost" $X \in \mathcal{L}^2$ a numerical surrogate $\mathcal{R}(X) \in (-\infty, \infty]$, what axioms?

Definition of coherency

$\mathcal{R}$ is a **coherent measure of risk** in the **basic** sense if

(R1) $\mathcal{R}(C) = C$ for all constants C

(R2) $\mathcal{R}((1-\lambda)X + \lambda X') \leq (1-\lambda)\mathcal{R}(X) + \lambda\mathcal{R}(X')$ for $\lambda \in (0,1)$ (convexity)

(R3) $\mathcal{R}(X) \leq \mathcal{R}(X')$ when $X \leq X'$ (monotonicity)

(R4) $\mathcal{R}(X) \leq c$ when $X_k \to X$ with $\mathcal{R}(X_k) \leq c$ (closedness)

(R5) $\mathcal{R}(\lambda X) = \lambda\mathcal{R}(X)$ for $\lambda > 0$ (positive homogeneity)

$\mathcal{R}$ is a coherent measure of risk in the **extended** sense when it satisfies (R1)–(R4), but not necessarily (R5)

(from ideas of Artzner, Delbaen, Eber, Heath 1997/1999)

(R1)+(R2) $\Rightarrow$ $\mathcal{R}(X + C) = \mathcal{R}(X) + C$ for all X and constants C

(R2)+(R5) $\Rightarrow$ $\mathcal{R}(X + X') \leq \mathcal{R}(X) + \mathcal{R}(X')$ (subadditivity)

Associated Criteria for Risk Acceptability

For a "cost" random variable X, to what extent should outcomes $X(\omega) > 0$, in constrast to outcomes $X(\omega) \leq 0$, be tolerated?

There is no single answer—this has to depend on **preferences!**

Preference-based definition of acceptance

Given a choice of a risk measure $\mathcal{R}$:

the risk in X is deemed acceptable when $\mathcal{R}(X) \leq 0$

(examples to come will illuminate this concept of Artzner et al.)

Notes:

from (R1): $\mathcal{R}(X) \leq c \iff \mathcal{R}(X - c) \leq 0$

from (R3): $\mathcal{R}(X) \leq \sup X$ for all X,

so X is always acceptable when $\sup X \leq 0$

(i.e., when there is **no chance** of an outcome $X(\omega) > 0$)

Consequences of Coherency for Optimization

For $i = 0, 1, \ldots, m$ let $\mathcal{R}_i$ be a coherent measure of risk in the **basic** sense, and consider the reconstituted problem:

minimize $\bar{c}_0(x)$ over $x \in S$ with $\bar{c}_i(x) \leq 0$ for $i = 1, \ldots, m$

where $\bar{c}_i(x) = \mathcal{R}_i(\underline{c}_i(x))$ for $\underline{c}_i(x) : \omega \to c_i(x, \omega)$

Key properties

(a) (preservation of convexity) If $c_i(x, \omega)$ is convex with respect to x, then the same is true for $\bar{c}_i(x)$

(so convex programming models persist)

(b) (preservation of certainty) If $c_i(x, \omega)$ is a value $c_i(x)$ independent of ω, then $\bar{c}_i(x)$ is that same value

(so features not subject to uncertainty are left undistorted)

(c) (insensitivity to scaling) The optimization problem is unaffected by rescaling of the units of the c_i's.

(a) and (b) still hold for coherent measures in the extended sense

Coherency or Its Lack in Traditional Approaches

Assessing the risk in each $\underline{c}_i(x)$ as $\mathcal{R}_i(\underline{c}_i(x))$ for a choice of $\mathcal{R}_i$

The case of Approach 1: guessing the future

$\mathcal{R}_i(X) = X(\bar{\omega})$ for a choice of $\bar{\omega} \in \Omega$ with prob > 0

$\mathcal{R}_i$ is **coherent**—but open to criticism

$\underline{c}_i(x)$ is deemed to be risk-acceptable if merely $c_i(x, \bar{\omega}) \leq 0$

The case of Approach 2: worst case analysis

$\mathcal{R}_i(X) = \sup X$

$\mathcal{R}_i$ is **coherent**—but very conservative

$\underline{c}_i(x)$ is risk-acceptable only if $c_i(x, \omega) \leq 0$ with prob $= 1$

The case of Approach 3: relying on expectations

$\mathcal{R}_i(X) = \mu(X) = EX$

$\mathcal{R}_i$ is **coherent**—but perhaps too “feeble”

$\underline{c}_i(x)$ is risk-acceptable as long as $c_i(x, \omega) \leq 0$ on average

The case of Approach 4: standard deviation units as safety margins

$\mathcal{R}_i(X) = \mu(X) + \lambda_i \sigma(X)$ for some $\lambda_i > 0$

$\mathcal{R}_i$ is **not coherent**: the monotonicity axiom (R3) fails!

$\Longrightarrow$ $\underline{c}_i(x)$ could be deemed more costly than $\underline{c}_i(x')$ even though $c_i(x,\omega) < c_i(x',\omega)$ with probability 1

$\underline{c}_i(x)$ is risk-acceptable as long as the mean $\mu(\underline{c}_i(x))$ lies below 0 by at least λ_i times the amount $\sigma(\underline{c}_i(x))$

The case of Approach 5: specifying probabilities of compliance

$\mathcal{R}_i(X) = q_{\alpha_i}(X)$ for some $\alpha_i \in (0,1)$, where

$q_{\alpha_i}(X) = \alpha_i$-quantile in the distribution of X

(to be explained)

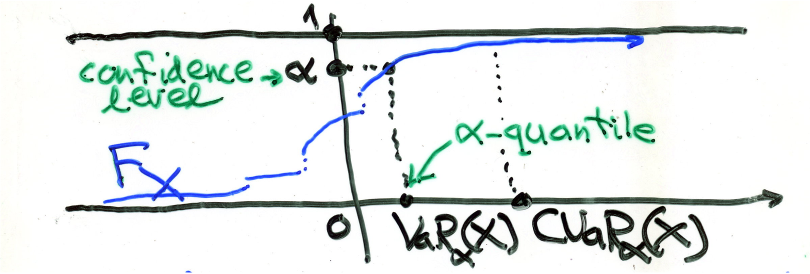
$\mathcal{R}_i$ is **not coherent**: the convexity axiom (R2) fails!

$\Longrightarrow$ for portfolios, this could run counter to "diversification"

$\underline{c}_i(x)$ is risk-acceptable as long as $c_i(x,\omega) \le 0$ with prob $\ge \alpha_i$

What further alternatives, remedies?

Quantiles and Conditional Value-at-Risk



α-quantile for X: $q_\alpha(X) = \min\{z \mid F_X(z) \geq \alpha\}$

value-at-risk: $\mathrm{VaR}_\alpha(X)$ same as $q_\alpha(X)$

conditional value-at-risk: $\mathrm{CVaR}_\alpha(X) =$ α-tail expectation of X

$$= \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\beta(X)d\beta \;\geq\; \mathrm{VaR}_\alpha(X)$$

THEOREM $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$ is a **coherent** measure of risk!

$\mathrm{CVaR}_\alpha(X) \nearrow \sup X$ as $\alpha \nearrow 1$, $\mathrm{CVaR}_\alpha(X) \searrow EX$ as $\alpha \searrow 0$

CVaR Versus VaR in Modeling

$$P\{X \le 0\} \le \alpha \iff q_\alpha(X) \le 0 \iff \mathrm{VaR}_\alpha(X) \le 0$$

Approach 5 recast: specifying probabilities of compliance

- focus on value-at-risk for the random variables $\underline{c}_i(x)$
- minimize $\mathrm{VaR}_{\alpha_0}(\underline{c}_0(x))$ over $x \in S$ subject to
  $\mathrm{VaR}_{\alpha_i}(\underline{c}_i(x)) \le 0, \; i = 1, \dots, m$
- pro/con: seemingly natural, but "incoherent" in general

Approach 6: safeguarding with conditional value-at-risk

- conditional value-at-risk instead of value-at-risk for each $\underline{c}_i(x)$
- minimize $\mathrm{CVaR}_{\alpha_0}(\underline{c}_0(x))$ over $x \in S$ subject to
  $\mathrm{CVaR}_{\alpha_i}(\underline{c}_i(x)) \le 0, \; i = 1, \dots, m$
- pro/con: coherent! also more cautious than value-at-risk

extreme cases: "$\alpha_i = 0$" $\sim$ expectation, "$\alpha_i = 1$" $\sim$ supremum

Some Elementary Portfolio Examples

securities $j = 1, \ldots, n$ with rates of return $\underline{r}_j$ and weights x_j

$$S = \{x = (x_1, \ldots, x_n) \,|\, x_j \geq 0,\ x_1 + \cdots + x_n = 1\}$$

rate of return of x-portfolio: $\underline{r}(x) = -[x_1 \underline{r}_1 + \cdots + x_n \underline{r}_n]$

$\underline{c}_0(x) = -\underline{r}(x), \quad \underline{c}_1(x) = \underline{q} - \underline{r}(x)$ with $\underline{q} \equiv -0.04$ here

Problems 1(a)(b)(c): expectation objective, CVaR constraints

(a) minimize $E[\underline{c}_0(x)]$ over $x \in S$
(b) minimize $E[\underline{c}_0(x)]$ over $x \in S$ subject to $\mathrm{CVaR}_{0.8}(\underline{c}_1(x)) \leq 0$
(c) minimize $E[\underline{c}_0(x)]$ over $x \in S$ subject to $\mathrm{CVaR}_{0.9}(\underline{c}_1(x)) \leq 0$

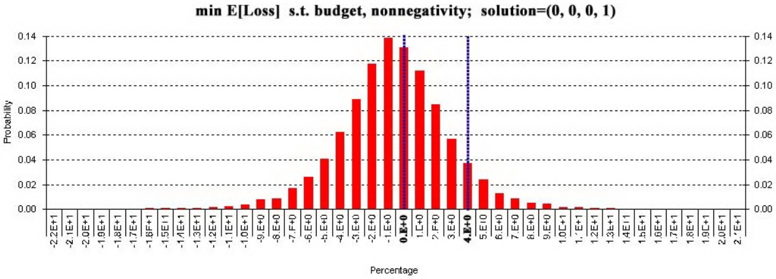
Problems 2(a)(b)(c): CVaR objectives, no benchmark constraints

(a) minimize $E[\underline{c}_0(x)]$ over $x \in S$ $\quad E[\underline{c}_0(x)] = \mathrm{CVaR}_{0.0}(\underline{c}_0(x))$
(b) minimize $\mathrm{CVaR}_{0.8}(\underline{c}_0(x))$ over $x \in S$
(c) minimize $\mathrm{CVaR}_{0.9}(\underline{c}_0(x))$ over $x \in S$

Portfolio Rate-of-Loss Contours, Problems 1(a)(b)(c)

Solutions computed with *Portfolio Safeguard* software, available for evaluation from American Optimal Decisions www.AOrDa.com

Results for Problem 1(a)

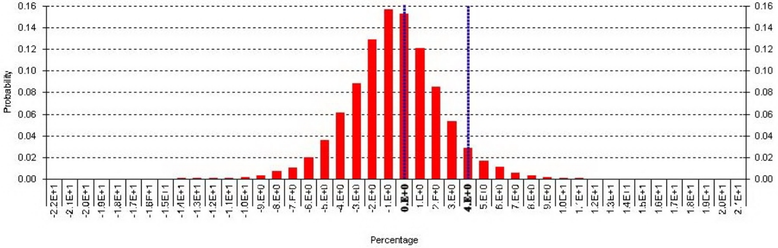


Solution vector: the portfolio weights for four different stocks

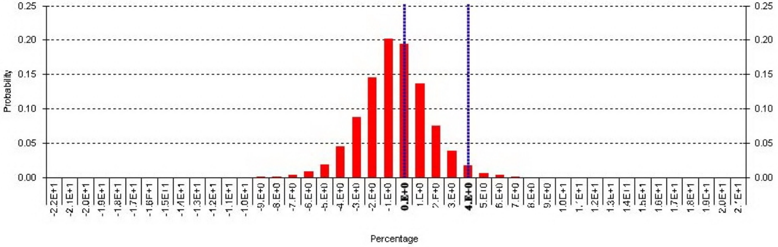
Note that in this case all the weight goes to the risky fourth stock

Results for Problems 1(b) and 1(c)

min E[Loss] s.t. CVaR{80%}(Loss) <=0.04, budget, nonnegativity; solution=(0.17, 0.04, 0.18, 0.61)



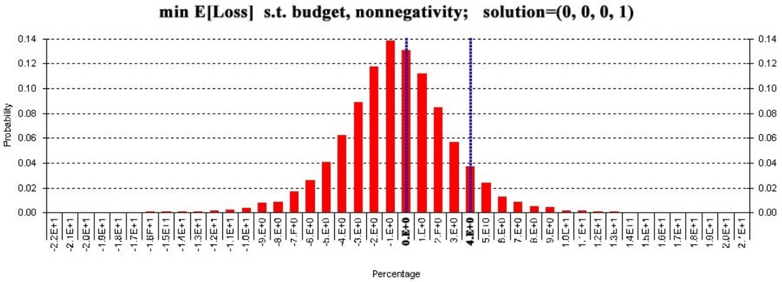
min E[Loss] s.t. CVaR{90%}(Loss) <=0.04, budget, nonnegativity; solution=(0.44, 0.40, 0.09, 0.07)



Portfolio Rate-of-Loss Contours, Problems 2(a)(b)(c)

Solutions computed with *Portfolio Safeguard* software, available for evaluation from American Optimal Decisions www.AOrDa.com

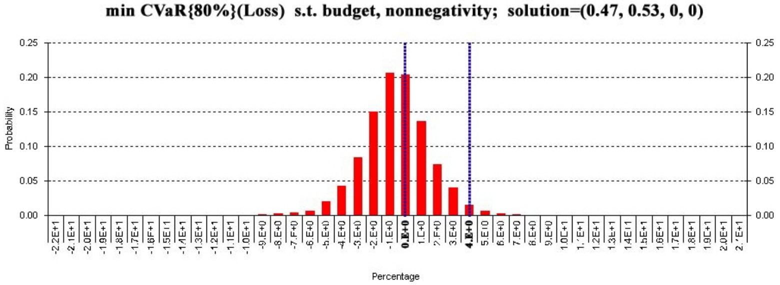
Results for Problem 2(a), same as Problem 1(a)

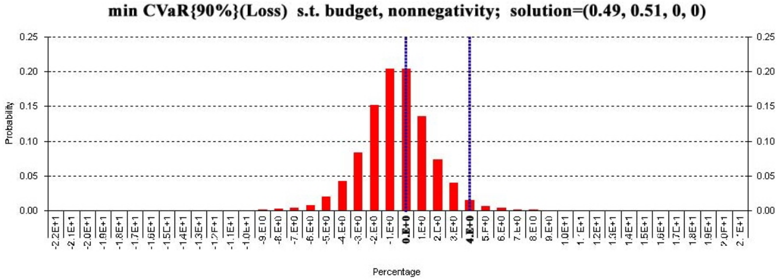


Solution vector: the portfolio weights for four different stocks

Again, in this case all the weight goes to the risky fourth stock

Results for Problems 2(b) and 2(c)





Minimization Formula for VaR and CVaR

$$\mathrm{CVaR}_\alpha(X) = \min_{C\in R}\Big\{ C + \frac{1}{1-\alpha} E\Big[\max\{0, X - C\}\Big]\Big\}$$

$$\mathrm{VaR}_\alpha(X) = \text{lowest } C \text{ in the interval giving the min}$$

min values behave better parametrically than minimizing points!

Application to CVaR optimization: convert a problem like

minimize $\mathrm{CVaR}_{\alpha_0}(\underline{c}_0(x))$ over $x \in S$ subject to

$$\mathrm{CVaR}_{\alpha_i}(\underline{c}_i(x)) \leq 0,\ i = 1,\dots,m$$

into a problem for $x \in S$ and auxiliary variables $C_0, C_1, \dots, C_m$:

minimize $C_0 + \frac{1}{1-\alpha_0} E\Big[\max\{0, \underline{c}_0(x) - C_0\}\Big]$ while requiring

$$C_i + \frac{1}{1-\alpha_i} E\Big[\max\{0, \underline{c}_i(x) - C_i\}\Big] \leq 0,\quad i = 1,\dots,m$$

Important case: this converts to **linear programming** when

(1) each $c_i(x,\omega)$ depends linearly on x,

(2) the future state space Ω is finite

(as is common in financial modeling, for instance)

Further Modeling Possibilities

additional sources of coherent measures of risk

Coherency-preserving combinations of risk measures

(a) If $\mathcal{R}_1, \ldots, \mathcal{R}_r$ are coherent and $\lambda_1 > 0, \ldots, \lambda_r > 0$ with $\lambda_1 + \cdots + \lambda_r = 1$, then

$$\mathcal{R}(X) = \lambda_1 \mathcal{R}_1(X) + \cdots + \lambda_r \mathcal{R}_r(X) \text{ is coherent}$$

(b) If $\mathcal{R}_1, \ldots, \mathcal{R}_r$ are coherent, then

$$\mathcal{R}(X) = \max\left\{\mathcal{R}_1(X), \ldots, \mathcal{R}_r(X)\right\} \text{ is coherent}$$

Example: $\mathcal{R}(X) = \lambda_1 \mathrm{CVaR}_{\alpha_1}(X) + \cdots + \lambda_r \mathrm{CVaR}_{\alpha_r}(X)$

Approach 7: safeguarding with CVaR mixtures

The CVaR approach already considered can be extended by replacing single CVaR expressions with weighted combinations

Continuous CVaR Mixtures and Risk Profiles

For any nonnegative **weighting** measure λ on $(0,1)$, a coherent measure of risk (in the basic sense) is given by

$$\mathcal{R}(X) = \int_0^1 \mathrm{CVaR}_\alpha(X)\, d\lambda(\alpha)$$

Spectral representation

Associate with λ the **profile** function. $\varphi(\alpha) = \int_0^\alpha [1-\beta]^{-1}\, d\lambda(\beta)$
Then, as long as $\varphi(1) < \infty$, the above $\mathcal{R}$ has the expression

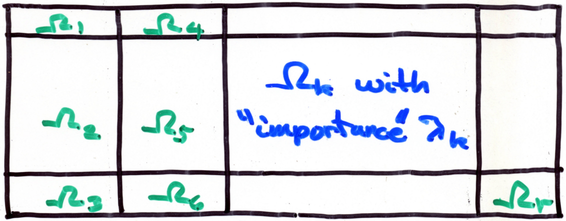
$$\mathcal{R}(X) = \int_0^1 \mathrm{VaR}_\beta(X)\varphi(\beta)\, d\beta$$

The functions φ arising this way as profiles are the nondecreasing, right-continuous functions $\varphi : [0,1] \to \boldsymbol{R}$ with $\varphi(0) = 0$

finite discrete mixtures correspond to *step functions* φ

Risk Measures From Subdividing the Future

“robust” optimization modeling revisited with Ω subdivided



$\lambda_k > 0$ for $k = 1, \ldots, r, \quad \lambda_1 + \cdots + \lambda_r = 1$

$\mathcal{R}(X) = \lambda_1 \sup_{\omega \in \Omega_1} X(\omega) \cdots + \lambda_r \sup_{\omega \in \Omega_r} X(\omega)$ is **coherent**

Approach 8: distributed worst-case analysis

Extend the ordinary worst-case model

$$\text{minimize } \sup_{\omega \in \Omega} c_0(x, \omega) \text{ subject to } \sup_{\omega \in \Omega} c_i(x, \omega) \leq 0 \, , i = 1, \ldots, m$$

by **distributing** each supremum **over subregions** of Ω, as above

Risk Envelope Characterization of Coherency

for coherent risk measures in the **basic** sense

A subset $\mathcal{Q}$ of $\mathcal{L}^2$ is a **coherent risk envelope** if it is nonempty, closed and convex, and $Q \in \mathcal{Q} \implies Q \geq 0,\ EQ = 1$

Interpretation: Any such Q is the "density" relative to the probability measure P on Ω of an alternative probability measure P' on Ω : $E_{P'}[X] = E[XQ],\ Q = dP'/dP$

[specifying $\mathcal{Q}$] $\longleftrightarrow$ [specifying a comparison set of measures P']

THEOREM: There is a one-to-one correspondence $\mathcal{R} \leftrightarrow \mathcal{Q}$ between coherent measures of risk $\mathcal{R}$ (in the basic sense) and coherent risk envelopes Q, which is furnished by the relations

$$\mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} E[XQ], \qquad \mathcal{Q} = \{ Q \,|\, E[XQ] \leq \mathcal{R}(X) \text{ for all } X \}$$

Examples and Extensions

$$\mathcal{R}(X) = EX \;\leftrightarrow\; \mathcal{Q} = \{1\}$$

$$\mathcal{R}(X) = \sup X \;\leftrightarrow\; \mathcal{Q} = \left\{ \text{all } Q \geq 0,\; EQ = 1 \right\}$$

$$\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X) \;\leftrightarrow\; \mathcal{Q} = \left\{ Q \geq 0,\; EQ = 1,\; Q \leq (1-\alpha)^{-1} \right\}$$

For coherent risk measures in the **extended** sense (not positively homogeneous) the corresponding representation is

$$\mathcal{R}(X) = \sup_Q \left\{ E[XQ] - \mathcal{I}(Q) \right\}, \quad \mathcal{I} = \mathcal{R}^*$$

where $\mathcal{I}$ is an lsc convex functional such that

$$\mathrm{cl}(\mathrm{dom}\,\mathcal{I}) \text{ is a risk envelope } \mathcal{Q} \text{ and } \min \mathcal{I} = 0 = \mathcal{I}(1)$$

Example: $\mathcal{R}(X) = \log E\left\{ e^X \right\} \;\leftrightarrow\; \mathcal{I}(Q) = E\left\{ Q \log Q \right\}$

Some References

[1] R. T. Rockafellar (2007), "Coherent approaches to risk in optimization under uncertainty," Tutorials in Operations Research INFORMS 2007, 39–61.

[2] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath (1999), "Coherent measures of risk," *Mathematical Finance* 9, 203–227.

[3] H. Föllmer, A. Schied (2002, 2004), *Stochastic Finance.*

[4] R.T. Rockafellar, S.P. Uryasev (2000), " Optimization of Conditional Value-at-Risk," *Journal of Risk* 2, 21–42.

[5] R.T. Rockafellar, S.P. Uryasev,, "Conditional value-at-risk for general loss distributions," *Journal of Banking and Finance* 26, 1443–1471.

[1], [4], [5], downloadable:

www.math.washington.edu/∼rtr/mypage.html

DEVIATION MEASURES AND GENERALIZED LINEAR REGRESSION

Terry Rockafellar
University of Washington, Seattle
University of Florida, Gainesville

Humboldt University, Berlin —— January, 2009

LECTURE 3

Quantification of Uncertainty

Framework for random variables X as before: $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$

orientation: $X(\omega)$ stands for a "cost" or loss

Axioms for deviation from constancy

$\mathcal{D}$ is a **measure of deviation** in the **basic** sense if

(D1) $\mathcal{D}(X) = 0$ for $X \equiv C$ constant, $\mathcal{D}(X) > 0$ otherwise

(D2) $\mathcal{D}((1-\lambda)X + \lambda X') \leq (1-\lambda)\mathcal{D}(X) + \lambda \mathcal{D}(X')$ for $\lambda \in (0,1)$ (convexity)

(D3) $\mathcal{D}(X) \leq c$ when $X_k \to X$ with $\mathcal{D}(X_k) \leq c$ (closedness)

(D4) $\mathcal{D}(\lambda X) = \lambda \mathcal{D}(X)$ for $\lambda > 0$ (positive homogeneity)

It is a **coherent** measure of deviation if it also satisfies

(D5) $\mathcal{D}(X) \leq \sup X - EX$ for all X

Deviation measures in the **extended** sense: (D4) dropped

$\Longrightarrow$ $\mathcal{D}$ actually has $\mathcal{D}(X + C) = \mathcal{D}(X)$ for all constants C

Initial Examples of Deviation Measures

notation: $X = X_+ - X_-$ for $X_+ = \max\{X, 0\}$, $X_- = \max\{-X, 0\}$

Standard deviation and semideviations

- $\sigma(X) = ||X - EX||_2$
- $\sigma_+(X) = ||[X - EX]_+||_2$ and $\sigma_-(X) = ||[X - EX]_-||_2$

Range-based deviation measures

- $\mathcal{D}(X) = \sup X - \inf X$
- $\mathcal{D}(X) = \sup X - EX$ and $\mathcal{D}(X) = EX - \inf X$

Recall that the $\mathcal{L}^p$ norms on $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ are well defined

$\mathcal{L}^p$ deviations and semideviations

- $\mathcal{D}(X) = ||X - EX||_p$
- $\mathcal{D}(X) = ||[X - EX]_+||_p$ and $\mathcal{D}(X) = ||[X - EX]_-||_p$

Risk Measures Paired With Deviation Measures

$\mathcal{R}$ is an **averse** measure of risk if it satisfies (R1), (R2), (R4) and

(R6) $\mathcal{R}(X) > EX$ for all nonconstant X (aversity)

basic sense: homogeneity (R5) yes, **extended** sense: (R5) no

Note: monotonicity axiom (R3) relinquished for this purpose

deviation measures versus risk measures

A one-to-one correspondence $\mathcal{D} \leftrightarrow \mathcal{R}$ between deviation measures $\mathcal{D}$ and **averse** measures $\mathcal{R}$ is furnished by

$$\mathcal{R}(X) = EX + \mathcal{D}(X), \qquad \mathcal{D}(X) = \mathcal{R}(X - EX)$$

and moreover $\mathcal{R}$ is coherent $\iff$ $\mathcal{D}$ is coherent

Example of CVaR deviation measures

- $\mathcal{D}(X) = \mathrm{CVaR}_\alpha(X - EX)$ is coherent
- $\mathcal{D}(X) = \int_0^1 \mathrm{CVaR}_\alpha(X - EX) d\lambda(\alpha)$ is coherent
  for any weighting measure λ on $(0, 1)$

Safety Margins Revisited

Recall the traditional approach to EX being "safely" below 0:

$EX + \lambda\sigma(X) \leq 0$ for some $\lambda > 0$ scaling the "safety"

but $\mathcal{R}(X) = EX + \lambda\sigma(X)$ is not **coherent**

Can the coherency be restored if $\sigma(X)$ is replaced by some $\mathcal{D}(X)$?

Yes! $\mathcal{R}(X) = EX + \lambda\mathcal{D}(X)$ is coherent when D is coherent

Safety margin modeling with coherency

In the safeguarding problem model

minimize $\bar{c}_0(x)$ over $x \in S$ with $\bar{c}_i(x) \leq 0$ for $i = 1, \ldots, m$

where $\bar{c}_i(x) = \mathcal{R}_i(\underline{c}_i(x))$ for $\underline{c}_i(x) : \omega \to c_i(x, \omega)$

coherency is obtained with

$\mathcal{R}_i(X) = EX + \lambda_i\mathcal{D}_i(X)$ for $\lambda_i > 0$ and $\mathcal{D}_i$ coherent

Generalized Deviations in Portfolio Optimization

financial instruments $i = 0, 1, \ldots, m$ with rates of return r_i

r_0 fixed, $r_1, \ldots, r_m$ random variables

Portfolio: given by "weights" $x_0, x_1, \ldots, x_m$, yielding $\sum_{i=0}^m x_i r_i$

Fundamental problem, generalized

minimize $\mathcal{D}(-\sum_{i=0}^m x_i r_i)$ for $\sum_{i=0}^m x_i = 1$, $E\{\sum_{i=0}^m x_i r_i\} = r_0 + \Delta$

Substituting $x_0 = 1 - x_1 - \cdots - x_m$ makes

$$x_0 r_0 + \sum_{i=1}^m x_i r_i = r_0 + \sum_{i=1}^m x_i [r_i - r_0]$$

Reformulations of the problem

In terms of $Y(x) = Y(x_1, \ldots, x_m) = -\sum_{i=1}^m x_i [r_i - r_0]$

minimize $\mathcal{D}(Y(x))$ over all $x \in R^n$ with $E[Y(x)] = -\Delta$

or for the associated risk measure $\mathcal{R}(X) = EX + \mathcal{D}(X)$

minimize $\mathcal{R}(Y(x))$ over all $x \in R^n$ with $E[Y(x)] = -\Delta$

Linear Regression

Approximation of a random variable Y by a linear combination of other random variables $X_1, \ldots, X_n$ and a constant term:

$$Y \;\approx\; c_0 + c_1 X_1 + c_2 X_2 + \cdots + c_n X_n$$

- Classical regression ...
- Quantile regression ...
- Other approaches? Why?

Should "risk preferences" dictate the form of approximation?

Underestimates worse than overestimates for $Y = \text{loss}/\text{cost}$!

Quantification of Error in Approximation

orientation: $X(\omega)$ refers to an outcome desired to be 0

Error measures $\mathcal{E} : \mathcal{L}^2 \to [0, \infty]$

$\mathcal{E}(X)$ quantifies the overall "nonzero-ness" in X

Error axioms

$\mathcal{E}$ is a **measure of error** in the **basic** sense if

(E1) $\mathcal{E}(0) = 0$, $\mathcal{E}(X) > 0$ when $X \neq 0$,
$\mathcal{E}(C) < \infty$ for all constants C

(E2) $\mathcal{E}((1-\lambda)X + \lambda X') \leq (1-\lambda)\mathcal{E}(X) + \lambda \mathcal{E}(X')$
for $\lambda \in (0,1)$ (convexity)

(E3) $\mathcal{E}(X) \leq c$ when $X_k \to X$ with $\mathcal{E}(X_k) \leq c$ (closedness)

(E4) $\exists\, \delta > 0$ with $\mathcal{E}(X) \geq \delta |EX|$ for all X (nondegeneracy)

(E5) $\mathcal{E}(\lambda X) = \lambda \mathcal{E}(X)$ for $\lambda > 0$ (positive homogeneity)

Error measures in the **extended** sense: (E5) dropped

Note: the nondegeneracy in (E4) is automatic in finite dimensions

Some Examples of Error Measures

$\mathcal{E} : \mathcal{L}^2 \to [0, \infty]$, basic if positively homogeneous

A broad class of error messages in the basic sense

$$\mathcal{E}(X) = ||a[X]_+ + b[X]_-||_p \text{ with } a > 0,\ b > 0,\ p \in [1, \infty]$$

Some specific instances:

$\mathcal{E}(X) = ||X||_p$ for $a = 1$ and $b = 1$

$\mathcal{E}(X) = E\{(1-\alpha)^{-1}X_+ - X\}$ for $a = (1-\alpha)^{-1}$, $b = 1$

Koenker-Basset error relative to $\alpha \in (0, 1)$

Generalized Regression

Let $Y, X_1, \ldots, X_n$ be random variables in $\mathcal{L}^2$

assume no linear combination of $X_1, \ldots, X_n$ is constant

Regession problem

For a measure $\mathcal{E}$ of error in the basic sense, with $\mathcal{E}(Y) < \infty$, choose $c_0, c_1, \ldots, c_n$ to

$$\text{minimize } \mathcal{E}\{Y - [c_0 + c_1 X_1 + \cdots + c_n X_n]\}$$

minimizing a **convex** function of $(c_0, c_1, \ldots, c_n) \in \boldsymbol{R}^{n+1}$

Existence of solutions

Optimal regression coefficient vectors $(\bar{c}_0, \bar{c}_1, \ldots, \bar{c}_n)$ exist and they form a compact convex set: $\mathcal{C}(Y) \subset \boldsymbol{R}^{n+1}$

Observe through axiom E5: $\mathcal{C}(\lambda Y) = \lambda \mathcal{C}(Y)$ for $\lambda > 0$

Portfolio Motivation

$Y_1, \ldots, Y_m$ = rates of return of various instruments
$x_1, \ldots, x_m$ = weights of these instruments in portfolio
$Y(x_1, \ldots, x_m) = x_1 Y_1 + \cdots + x_m Y_m$ = portfolio rate of return

Optimization context

Minimize some $\mathcal{R}$ or $\mathcal{D}$ aspect of $Y(x_1, \ldots, x_m)$ under some constraints on various other $\mathcal{R}$ or $\mathcal{D}$ aspects

Factor models

Simplication via "factors" $X_1, \ldots, X_n$:
each Y_i approximated by $\hat{Y}_i = c_{i0} + c_{i1} X_1 + \cdots + c_{in} X_n$
$Y(x_1, \ldots, x_m)$ replaced in optimization by $\hat{Y}(x_1, \ldots, x_m)$

Should the "risks" under consideration influence the approach taken to regression? Different regression for different $\mathcal{R}$ or $\mathcal{D}$?

Error Projection

$\mathcal{E}$ = any measure of error in the basic sense

Deviation measures from error measures

In terms of constants $C \in \boldsymbol{R}$, let

$$\mathcal{D}(X) = \inf_C \mathcal{E}(X - C), \qquad \mathcal{S}(X) = \operatorname*{argmin}_C \mathcal{E}(X - C)$$

- $\mathcal{D}$ is a deviation measure in the basic sense
- $\mathcal{S}(X)$ is, for every X, a nonempty closed interval in $\boldsymbol{R}$

(reducing typically to a single value, but not always)

$\mathcal{S}(X)$ is the associated "**statistic**"

Classical regression ("least squares")

$$\mathcal{E}(X) = \lambda||X||_2 \text{ for some } \lambda > 0$$
$$\mathcal{S}(X) = \mu(X) = EX$$
$$\mathcal{D}(X) = \lambda\sigma(X)$$

Nonclassical Examples of Regression

Regression with range deviation

$$\mathcal{E}(X) = \lambda||X||_\infty \text{ for some } \lambda > 0$$
$$\mathcal{S}(X) = \tfrac{1}{2}[\sup X + \inf X] \quad \text{center of range}$$
$$\mathcal{D}(X) = \tfrac{\lambda}{2}[\sup X - \inf X] \quad \text{radius of range, scaled}$$

Regression with mean absolute deviation

$$\mathcal{E}(X) = \lambda||X||_1 = \lambda E|X| \text{ for some } \lambda > 0$$
$$\mathcal{S}(X) = \operatorname{med} X \quad \text{median}$$
$$\mathcal{D}(X) = \lambda E[\operatorname{dist}(X, \operatorname{med} X)]$$

Note that $\operatorname{med} X = [\operatorname{med}^- X, \operatorname{med}^+ X]$, is an interval in general!

$$\mathcal{D}(X) = \lambda E[X - \operatorname{med} X] \text{ when } \operatorname{med}^- X = \operatorname{med}^+ X$$

Quantiles and Quantile Regression

recall: F_X = c.d.f. for X, $F_X(x) = P(X \leq x)$

Quantile interval for $\alpha \in (0,1)$:

$$q_\alpha(X) = [q_\alpha^-(X), q_\alpha^+(X)], \text{ where}$$

$$q_\alpha^-(X) = \inf\{x \mid F_X(x) \geq \alpha\},$$
$$q_\alpha^+(X) = \sup\{x \mid F_X(x) \leq \alpha\}$$

Quantile regression

$\mathcal{E}(X) = E\{\,(1-\alpha)^{-1}[X]^+ - X\,\}$ Koenker-Basset error

$\mathcal{S}(X) = q_\alpha(X)$ α-quantile

$\mathcal{D}(X) = \mathrm{CVaR}_\alpha(X - EX)$

Regression Analysis

Approximation goal: $Y \approx c_0 + c_1X_1 + \cdots + c_nX_n$

$$Z(c_0, c_1, \ldots, c_n) = Y - [c_0 + c_1X_1 + \cdots + c_nX_n]$$
$$Z_0(c_1, \ldots, c_n) = Y - [c_1X_1 + \cdots + c_nX_n]$$

REGRESSION PROBLEM for error measure $\mathcal{E}$:

minimize $\mathcal{E}(\, Z(c_0, c_1, \ldots, c_n)\,)$ over $c_0, c_1, \ldots, c_n$

THEOREM Error-shaping decomposition

The coefficients $\bar{c}_0, \bar{c}_1, \ldots, \bar{c}_n$ are optimal if and only if

$$(\bar{c}_1, \ldots, \bar{c}_n) \in \underset{c_1,\ldots,c_n}{\operatorname{argmin}}\, \mathcal{D}(\, Z_0(c_1, \ldots, c_n)\,)$$
$$\bar{c}_0 \in \mathcal{S}(\, Z_0(c_1, \ldots, c_n)\,)$$

COROLLARY Equivalent interpretation of regression

Choose $(c_0, c_1, \ldots c_n)$ to minimize $\mathcal{D}(Z(c_0, c_1, \ldots, c_n))$ subject to the requirement that $0 \in \mathcal{S}(Z(c_0, c_1, \ldots, c_n))$

Regression Interpreted in Examples

Approximation goal: $Y \approx c_0 + c_1 X_1 + \cdots + c_n X_n$

Regression error being shaped:

$$Z = Z(c_0, c_1, \ldots, c_n) = Y - [c_0 + c_1 X_1 + \cdots + c_n X_n]$$

1. Classical regression "least squares"
   minimize $\sigma(Z)$ subject to $EZ = 0$
2. Range regression
   minimize breadth of range of Z subject to center being 0
3. Median regression
   minimize $E|Z|$ subject to "median of Z being 0"
4. Quantile regression at quantile level $\alpha \in (0,1)$
   minimize $E[\,(1-\alpha)^{-1}|Z|^{+} - Z\,]$ subject to "$q_\alpha(Z) = 0$"
5. Mixed quantile regression . . . further illustrations

Portfolio Application

$Y_1, \ldots, Y_m =$ rates of return, $\quad x_1, \ldots, x_m =$ weights

Portfolio rate of return:

$$Y(x) = x_1 Y_1 + \cdots + x_m Y_m \text{ for } x = (x_1, \ldots, x_n)$$

Risk aspects of portfolio: in objective or constraints

$$f_{\mathcal{D}}(x) = \mathcal{D}(Y(x)) \text{ or } f_{\mathcal{R}}(x) = \mathcal{R}(Y(x)) \text{ for various } \mathcal{D}, \mathcal{R}$$

Factor model with factors $X_1, \ldots, X_n$:

$$Y_i \approx \hat{Y}_i(c_i) = c_{i0} + c_{i1} X_1 + \cdots + c_{in} X_n \text{ for each } i$$

$$Y(x) \approx \hat{Y}(x, c_1, \ldots, c_m) = x_1 \hat{Y}_1(c_1) + \cdots + x_m \hat{Y}_m(c_m)$$

Consequence for risk expressions:

$$f_{\mathcal{D}}(x) = \mathcal{D}(Y(x)) \approx \hat{f}_{\mathcal{D}}(x, c_1, \ldots, c_m) = \mathcal{D}(\hat{Y}(x, c_1, \ldots, c_m))$$

$$f_{\mathcal{R}}(x) = \mathcal{R}(Y(x)) \approx \hat{f}_{\mathcal{R}}(x, c_1, \ldots, c_m) = \mathcal{R}(\hat{Y}(x, c_1, \ldots, c_m))$$

How will these approximation errors affect optimization?

Complication: the errors must be treated parametrically in x!

Parametric Bounds: $\mathcal{D}$ Type

Factor approximation errors:

$$Z_i(c_{i0}, c_{i1}, \ldots, c_{in}) = Y_i - [c_{i0} + c_{i1}X_1 + \cdots + c_{in}X_n]$$

coefficient vectors $c_i = (c_{i0}, c_{i1}, \ldots, c_{in})$

Targeted inequality: with a coefficient vector $a \geq 0$

$$f_{\mathcal{D}}(x) \;\leq\; \hat{f}_{\mathcal{D}}(x, c_1, \ldots, c_m) + a \cdot x \quad \text{for all } x \geq 0$$

What is the "best" that can be achieved through the control of the factor approximation errors? lowest $a = (a_1, \ldots, a_n)$?

auxiliary notation: $Z_{i0}(c_{i1}, \ldots, c_{in}) = Y_i - [c_{i1}X_1 + \cdots + c_{in}X_n]$

THEOREM   The lowest $a = (a_1, \ldots, a_n)$ is achieved by

- determining $\bar{c}_i = (\bar{c}_{i0}, \bar{c}_{i1}, \ldots, \bar{c}_{in})$ through generalized regression using an error measure $\mathcal{E}$ that projects onto $\mathcal{D}$
- taking $a_i = \mathcal{D}(Z_{i0}(\bar{c}_{i1}, \ldots, \bar{c}_{in}))$ note: $\bar{c}_{i0}$ has no role

Parametric Bounds: $\mathcal{R}$ Type

Targeted inequality: with a coefficient vector $a \geq 0$

$$f_{\mathcal{R}}(x) \;\leq\; \hat{f}_{\mathcal{R}}(x, c_1, \ldots, c_m) + a \cdot x \quad \text{for all } x \geq 0$$

What is the "best" that can be achieved through the control of the factor approximation errors? lowest $a = (a_1, \ldots, a_n)$?

THEOREM The lowest $a = (a_1, \ldots, a_n)$ is achieved actually with $a = 0$! by

- determining $\bar{c}_i = (\bar{c}_{i0}, \bar{c}_{i1}, \ldots, \bar{c}_{in})$ through generalized regression using an error measure $\mathcal{E}$ that projects onto the deviation measure $\mathcal{D}$ corresponding to the risk measure $\mathcal{R}$
- replacing $\bar{c}_i$ by $\bar{c}_i^*$, with
  $\bar{c}_{i0}^* = \mathcal{R}(Z_{i0}(\bar{c}_{i1}, \ldots, \bar{c}_{in}))$, but $\bar{c}_{ij}^* = \bar{c}_{ij}$ for $j = 1, \ldots, n$.

Acceptability consequence:

$$\mathcal{R}(\hat{Y}(x, \bar{c}_1^*, \ldots, \bar{c}_m^*)) \leq 0 \quad \Longrightarrow \quad \mathcal{R}(Y(x)) \leq 0$$

Some References

[1] R.T. Rockafellar, S. Uryasev, M. Zabarankin (2006), "Generalized deviations in risk analysis," *Finance and Stochastics* 10, 51–74.

[2] R.T. Rockafellar, S. Uryasev, M. Zabarankin (2006), "Master funds in portfolio analysis with general deviation measures," *Journal of Banking and Finance* 30, 743–778.

[3] R.T. Rockafellar, S. Uryasev, M. Zabarankin (2006), "Optimality conditions in portfolio analysis with general deviation measures," *Math. Programming, Ser. B* 108, 515–540.

[4] R.T. Rockafellar, S. Uryasev, M. Zabarankin (2008), "Risk tuning in generalized linear regression," *Mathematics of Operations Research* 33, 712–729.

[5] R. Koenker, G. W. Bassett (1978), "Regression quantiles," *Econometrica* 46, 33–50.

UTILITY, GENERALIZED ENTROPY AND MEASURES OF LIABILITY

Terry Rockafellar
University of Washington, Seattle
University of Florida, Gainesville

Humboldt University, Berlin —— January, 2009

LECTURE 4

Integral Functionals

$(\Omega, \mathcal{F}, P)$ = some probability space

A closed-set-valued mapping $S : \Omega \to R^n$ is **measurable** when

$\{\omega \mid S(\omega) \cap C\} \in \mathcal{F}$ for all closed sets $C \subset R^n$

A function $f : R^n \times \Omega \to \bar{R}$ is a **normal integrand** when

$f(x, \omega)$ is lsc in x and $S : \omega \to \operatorname{epi} f(\cdot, \omega)$ is measurable

Consequence: $f(x(\omega), \omega)$ is measurable when $x(\omega)$ is measurable

Conjugacy on paired spaces $\mathcal{L}^p_n(\Omega, \mathcal{F}, P)$ and $\mathcal{L}^q_n(\Omega, \mathcal{F}, P)$

For a normal integrand f, the integral functional

$$I_f(x(\cdot) = E\{f(x(\cdot), \cdot)\} = \int_\Omega f(x(\omega), \omega) dP(\omega)$$

is (with minor assumption) well-defined for $x(\cdot) \in \mathcal{L}^p_n(\Omega, \mathcal{F}, P)$, and

$I_f^* = I_{f^*}$ on $\mathcal{L}^q_n(\Omega, \mathcal{F}, P)$, $\quad I_f^{**} = I_{f^{**}}$ on $\mathcal{L}^p_n(\Omega, \mathcal{F}, P)$

Note: I_f is **convex** when $f(x, \omega)$ is convex in x, and then

$$v(\cdot) \in \partial I_f(x(\cdot)) \iff v(\omega) \in \partial f(x(\omega), \omega) \text{ almost surely}$$

Utility Maximization in Finance

Instruments: $i = 0, 1, \ldots, m$ with returns X_i, risk-free for $i = 0$

prices π_i with $\pi_0 = 1$, rates of return $r_i = X_i/\pi_i - 1$, r_0 constant

$Y_i = X_i/[1 + r_0] - \pi_i$ gives net return in present money

Portfolios: weights ξ_i yielding $\sum_{i=0}^m \xi_i X_i$ at cost $\sum_{i=0}^m \xi_i \pi_i$, or

in present money yielding $\sum_{i=1}^m \xi_i Y_i + w$ from investment w

Monetary utility, normalized:

$u(x) =$ the amount of present money deemed acceptable

in lieu of receiving the future amount $[1 + r_0]x$

u is concave, nondecreasing, with $u(0) = 0$, $u(x) \leq x$

Utility maximization problem

maximize $E\{u(\sum_{i=1}^m \xi_i Y_i + w)\}$ over $\xi = (\xi_1, \ldots, \xi_m)$

$\mathcal{U}(X) = E\{u(X)\}$ assesses present worth of future gain $[1 + r_0]X$

Reformulation to Minimization in Loss Context

$v(x) = -u(-x) =$ the **liability** exposure associated with x

$=$ the amount of present money deemed necessary as compensation for losing $[1+r_0]x$ in the future

v is convex, nondecreasing, with $v(0) = 0$, $v(x) \geq x$

Liability minimization problem

minimize $E\{v(\sum_{i=1}^m \xi_i[-Y_i] - w)\}$ over $\xi = (\xi_1, \ldots, \xi_m)$

$\mathcal{V}(X) = E\{v(X)\} = I_v(X) =$ **integral** functional on $\mathcal{L}^p(\Omega, \mathcal{F}, P)$

$\mathcal{V}$ is convex, nondecreasing, with $\mathcal{V}(0) = 0$, $\mathcal{V}(X) \geq EX$

Conjugate: $\mathcal{V}^*(Q) = I_{v^*}(Q) = E\{v^*(Q)\}$ on $\mathcal{L}^q(\Omega, \mathcal{F}, P)$

$\mathcal{V}^*$ is convex, $\mathcal{V}^*(Q) \geq 0$, $\mathcal{V}^*(1) = 0$, and $\mathcal{V}^*(Q) < \infty \Rightarrow Q \geq 0$

Insurance interpretation: $\mathcal{V}(X)$ is the **premium** to be charged (relative to v) for covering the uncertain future loss $[1+r_0]X$

Lagrangian and Dual Problem

$$\mathcal{V}(X) = E\{v(X)\}, \qquad \mathcal{V}^*(Q) = E\{v^*(Q)\}$$

Lagrangian for the minimization problem:

$$L(\xi_1, \ldots, \xi_m; Q) = E\{(\sum_{i=1}^m \xi_i[-Y_i] + [-w])Q\} - \mathcal{V}^*(Q)$$

Derivation of the dual objective:

$$\begin{aligned} g(Q) &= \inf_{\xi_1,\ldots,\xi_m} L(\xi_1, \ldots, \xi_m; Q) \\ &= [-w]EQ - \mathcal{V}^*(Q) \text{ if } Q \geq 0 \text{ and } E[Y_i Q] = 0, \\ &\text{but } = -\infty \text{ otherwise} \end{aligned}$$

Dual problem

maximize $[-w]EQ - E\{v^*(Q)\}$ subject to
$Q \geq 0$ and $E[Y_i Q] = 0$ for $i = 1, \ldots, m$

$-w =$ the money extracted from the market in the present
for taking on the future losses associated with $\sum_{i=0}^m \xi_i[-X_i]$

Application of Duality Criteria

These primal and dual problems fit the extended Fenchel format:

$$(\mathcal{P}) \qquad \textit{minimize}\big\{\langle c,\xi\rangle + h(\xi) + k(b - A\xi)\big\},$$
$$(\mathcal{D}) \qquad \textit{maximize}\big\{\langle b,Q\rangle - k^*(Q) - h^*(A^*Q - c)\big\},$$

with $\xi \in \boldsymbol{R}^m$ and $Q \in \mathcal{L}^q$, paired with $\mathcal{L}^p$, $p < \infty$, by taking

$$c = 0,\ h \equiv 0,\ h^* = \delta_0,\ k = \mathcal{V},\ k^* = \mathcal{V}^*,\ b = -w,$$
$$A : \xi \to \sum_{i-1}^m \xi_i Y_i, \qquad A^* : Q \to (\,E[Y_1 Q], \ldots, E[Y_m Q]\,)$$

Criteria to be specialized:

$$b \in \operatorname{int}[A(\operatorname{dom} h) + \operatorname{dom} k], \qquad c \in \operatorname{int}[A^*(\operatorname{dom} k^*) - \operatorname{dom} h^*]$$

Duality theorem

(a) $\inf(\mathcal{P}) = \max(\mathcal{D})$ if $-w \in \operatorname{int}\big\{X \in \mathcal{L}^p \,\big|\, E\{v(X)\} < \infty\big\}$

(b) $\min(\mathcal{P}) = \sup(\mathcal{D})$ if

$$0 \in \operatorname{int}\big\{\,(\,E[Y_1 Q], \ldots, E[Y_m Q]\,)\,\big|\, Q \in \mathcal{L}^q,\ E\{v^*(Q)\} < \infty\big\}$$

It is possible also to work with $X \in \mathcal{L}^\infty$ and $Q \in (\mathcal{L}^\infty)^*$. Further analysis then relates the results to known conditions in finance.

Valuations of Liability Generalized

functionals $\mathcal{V}(X)$, not just of form $I_v(X)$, for potential losses X

Liability measures

Call $\mathcal{V}$ a **measure of liability** if: (V1) $\mathcal{V}(0) = 0$, $\mathcal{V}(X) \geq EX$, (V2) $\mathcal{V}$ convex, (V3) $\mathcal{V}$ nondecreasing, (V4) $\mathcal{V}$ lsc

Conjugate characterization:

$\mathcal{V}^*$ convex, lsc, $\mathcal{V}^*(Q) \geq 0$, $\mathcal{V}^*(1) = 0$, $\mathcal{V}^*(Q) < \infty \Rightarrow Q \geq 0$

Consider a **trade-off:** minimize $C + \mathcal{V}(X - C)$ over $C \in \mathbb{R}$

charge C up front, reducing uncertain future losses accordingly

Derivation of associated risk measure and entropy

(a) $\mathcal{R}(X) = \min_C \left\{ C + \mathcal{V}(X - C) \right\}$ is a coherent measure of risk

(b) $\mathcal{R}^*(Q) = \mathcal{V}^*(Q)$ if $EQ = 1$, but $\mathcal{R}^*(Q) = \infty$ otherwise

$\mathcal{R}^*(Q)$ is thus an **entropy** functional $\mathcal{I}(Q)$, $-\mathcal{I}(Q) =$ the entropy

Minimization of Portfolio Risk

$\mathcal{V}$ = measure of liability, $\mathcal{R}$ = associated risk, $\mathcal{I} = \mathcal{R}^*$ entropy

$$\mathcal{R}(\textstyle\sum_{i=1}^m \xi_i[-Y_i] - w) = \mathcal{R}(\sum_{i=1}^m \xi_i[-Y_i]) - w$$

Portfolio risk minimization problem

minimize $\mathcal{R}(\,\sum_{i=1}^m \xi_i[-Y_i]\,)$ over $\xi = (\xi_1, \ldots, \xi_m)$

Lagrangian function:

$$L(\xi_1, \ldots, \xi_m; Q) = E\{\,\textstyle\sum_{i=1}^m \xi_i[-Y_i]Q\} - \mathcal{I}(Q)$$
$$= \textstyle\sum_{i=1}^m \xi_i E\{[-Y_i]Q\} - \mathcal{V}^*(Q) \text{ if } Q \geq 0,\ EQ = 1$$

but $= -\infty$ otherwise

Corresponding dual problem in entropy

maximize $-\mathcal{I}(Q)$ subject to $E[Y_i Q] = 0$ for $i = 1, \ldots, m$

$\Rightarrow$ Q is a **risk neutral** probability density, $Q = dP^*/dP$

an "entropic distance" of P^* from the nominal P is minimized

Aversity in Liability Valuation

Call a liability measure $\mathcal{V}$ **averse** if $\mathcal{V}(X) > EX$ when $X \neq 0$

Associated measures of error and deviation

Let $\mathcal{V}$ be an averse measure of liability, and let $\mathcal{R}(X)$ be the associated measure of risk, $\mathcal{R}(X) = \min_C \{C + \mathcal{V}(X - C)\}$

(a) $\mathcal{R}(X)$ is an **averse measure of risk** and coherent
(b) $\mathcal{E}(X) = \mathcal{V}(X) - EX$ is a **measure of error**
(c) $\mathcal{D}(X) = \min_C \{\mathcal{E}(X - C)\}$ agrees with $\mathcal{D}(X) = \mathcal{R}(X - EX)$

Integral functional case: $\mathcal{V}(X) = E\{v(X)\}$

v convex, nondecreasing, with $v(0) = 0$, $v(x) \geq x$

$\mathcal{V}(X) = E\{v(X)\}$ is averse when $v(x) > x$ for $x \neq 0$
$\mathcal{E}(X) = E\{\varepsilon(X)\}$ for the function $\varepsilon(x) = v(x) - x$

CVaR Revisited

Consider the liability measure $\mathcal{V}(X) = E\{v(X)\}$ and associated error measure $\mathcal{E}(X) = E\{\varepsilon(X)\} = E\{v(X) - X\}$, deviation measure $\mathcal{D}(X) = \min_C \{\mathcal{E}(X - C)\}$ and coherent risk measure $\mathcal{R}(X) = \min_C \{C + \mathcal{V}(X - C)\}$ in the case of

$$v(x) = (1-\alpha)^{-1} \max\{x, 0\} \text{ (averse), with}$$
$$\varepsilon(x) = v(x) - x = [(1-\alpha)^{-1} - 1] \max\{x, 0\} + \max\{-x, 0\}$$

where $0 < \alpha < 1$, so that $(1-\alpha)^{-1} > 1$. Then

(a) $\mathcal{V}(X) = (1-\alpha)^{-1} E[X_+]$

(b) $\mathcal{E}(X) = [(1-\alpha)^{-1} - 1] E[X_+] + E[X_-]$ Koenker-Basset error

(c) $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$

(d) $\mathcal{D}(X) = \mathrm{CVaR}_\alpha(X - EX)$;

For "utility" version of this, see paper of Ben-Tal and Teboulle

Some References

[1] R. T. Rockafellar (1998), *Variational Analysis*, Springer-Verlag
Chapter 14 (for issues of measurability)

[2] R. T. Rockafellar (1971), "Integrals which are convex functionals, II," *Pacific Journal of Mathematics* 39, 439–469
conjugates on $(\mathcal{L}^\infty)*$ are covered as well

[3] R. T. Rockafellar (1976), "Integral functionals, normal integrands and measurable selections," in *Nonlinear Operators and the Calculus of Variations*, L. Waelbroeck (ed.), Lecture Notes in Math. 543, Springer-Verlag, 157-207
most now in [1], except weak compactness characterization

[4] A. Ben-Tal, M. Teboulle (2007), "An old-new concept of convex risk measures: the optimized certainty equivalent," *Mathematical Finance* 17, 449–476.