THE FUNDAMENTAL QUADRANGLE OF RISK

relating quantifications of various aspects of a random variable

---

$$\begin{array}{ccc} & \text{risk } \mathcal{R} \longleftrightarrow \mathcal{D} \text{ deviation} & \\ \text{optimization} & \uparrow \quad \mathcal{S} \quad \uparrow & \text{estimation} \\ & \text{regret } \mathcal{V} \longleftrightarrow \mathcal{E} \text{ error} & \end{array}$$

Lecture 1: optimization, the role of $\mathcal{R}$

Lecture 2: estimation, the roles of $\mathcal{E}$, $\mathcal{D}$, $\mathcal{S}$

Lecture 3: tying both together along with $\mathcal{V}$ and duality

Lecture 1

QUANTIFICATIONS OF RISK IN STOCHASTIC OPTIMIZATION

R. T. Rockafellar
University of Washington, Seattle
University of Florida, Gainesville

Newcastle, Australia **February, 2010**

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**?

The Fundamental Difficulty Caused by Uncertainty

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 $S \subset R^n$ and functions $c_i : S \mapsto R$

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: $\underline{c}_i(x) : \omega \mapsto c_i(x, \omega), \quad i = 0, 1, \ldots, m$

Example: Linear Programming Context

minimize $c_0(x)$ over all $x \in S$ satisfying $c_i(x) \leq 0,\ i = 1, \ldots, m$

Linear programming problem:

$$c_i(x) = a_{i1}x_1 + \cdots a_{in}x_n - b_i$$

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

Effect of uncertainty:

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

There is **no single clear answer** to the question of how then to reconstitute the optimization **objective** and the **constraints**!

Stochastic Framework — Random Variables

Future state space Ω modeled with a probability structure:

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

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

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

$$F_X(z) = \text{prob}\left\{\omega \,\middle|\, X(\omega) \leq z\right\}$$

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 $E[X^2] < \infty$

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** $\underline{c}_0(x), \underline{c}_1(x), \ldots, \underline{c}_m(x)$

Some Traditional Approaches

Recapturing optimization in the face of $\underline{c}_i(x) : \omega \to c_i(x,\omega)$

Approach 1: guessing the future

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

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

- 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 $x \in S$:

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

- 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 $x \in S$:

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

- 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 $\lambda_i > 0$

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

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

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

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

- 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

Traditional usage: problems of **reliable design** in engineering

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

The idea to be pursued here:

capture the "risk" in X by a **numerical surrogate** $\mathcal{R}(X)$

This leads to considering

functionals $\mathcal{R} : X \to \mathcal{R}(X)$ on the space of random variables

$\mathcal{R}$ = "risk quantifier" = "risk measure"

A Systematic Approach to Uncertainty in Optimization

When numerical values $c_i(x)$ become random variables $\underline{c}_i(x)$:

- choose risk quantifiers $\mathcal{R}_i$ for $i = 0, 1, \ldots, m$
- 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 $\bar{c}_i(x) \leq 0$, $i = 1, \ldots, m$.

Basic Guidelines

What axioms for numerical surrogates $\mathcal{R}(X) \in (-\infty, \infty]$?

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)

$\mathcal{R}$ coherent in the **extended** sense: axiom (R5) dropped

(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?

preferences must be articulated!

Preference-based definition of acceptance

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

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

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 $\sup X \leq 0$

The Role of Coherency in Optimization

Reconstituted optimization 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 $\underline{c}_i(x) : \omega \to c_i(x, \omega)$

Assumption for now: each $\mathcal{R}_i$ is **coherent** in the **basic** sense

Key properties associated with coherency

(a) (preservation of convexity)

$c_i(x, \omega)$ convex in $x \implies \bar{c}_i(x)$ convex in x

(b) (preservation of certainty)

$c_i(x, \omega)$ independent of $\omega \implies \bar{c}_i(x)$ has that same value

(c) (insensitivity to scaling)

optimization 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

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

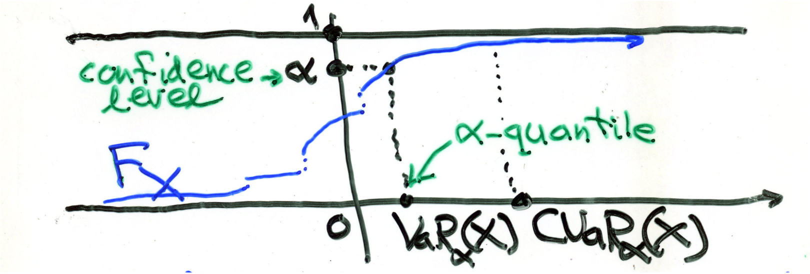
(to be explained)

$\mathcal{R}_i$ 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 $\ge \alpha_i$

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: $\mathrm{CVaR}_\alpha(X) = \alpha$-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

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

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
  $\mathrm{VaR}_{\alpha_i}(\underline{c}_i(x)) \le 0, \; i = 1, \ldots, m$
- 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
  $\mathrm{CVaR}_{\alpha_i}(\underline{c}_i(x)) \le 0, \; i = 1, \ldots, m$
- pro/con: coherent! also more cautious than value-at-risk

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

Minimization Formula for VaR and CVaR

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

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

Application to CVaR optimization: convert a problem like

$$\text{minimize } \mathrm{CVaR}_{\alpha_0}(\underline{c}_0(x)) \text{ over } x \in S \text{ subject to}$$
$$\mathrm{CVaR}_{\alpha_i}(\underline{c}_i(x)) \leq 0, \; i = 1, \ldots, m$$

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

$$\text{minimize } C_0 + \frac{1}{1-\alpha_0} E\left[\max\{0, \underline{c}_0(x) - C_0\} \right] \text{ while requiring}$$
$$C_i + \frac{1}{1-\alpha_i} E\left[\max\{0, \underline{c}_i(x) - C_i\} \right] \leq 0, \quad i = 1, \ldots, m$$

Further Modeling Possibilities

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

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

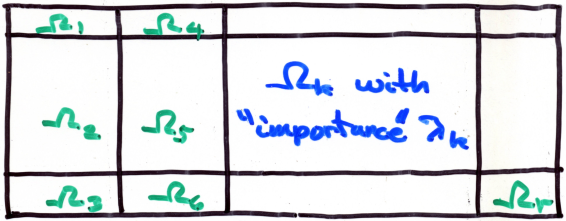
Spectral representation

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

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

Risk Measures From Subdividing the Future

"robust" optimization modeling revisited with Ω subdivided



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

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

Approach 8: distributed worst-case analysis

Extend the ordinary worst-case model

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

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

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

THE FUNDAMENTAL QUADRANGLE OF RISK

relating quantifications of various aspects of a random variable

---

$$\begin{array}{ccc} \text{risk } \mathcal{R} & \longleftrightarrow & \mathcal{D} \text{ deviation} \\ \uparrow & \mathcal{S} & \uparrow \\ \text{regret } \mathcal{V} & \longleftrightarrow & \mathcal{E} \text{ error} \end{array}$$

optimization — estimation

Lecture 1: optimization, the role of $\mathcal{R}$

Lecture 2: estimation, the roles of $\mathcal{E}$, $\mathcal{D}$, $\mathcal{S}$

Lecture 3: tying both together along with $\mathcal{V}$ and duality

Lecture 2

QUANTIFICATIONS OF ERROR IN GENERALIZED REGRESSION AND ESTIMATION

R. T. Rockafellar
University of Washington, Seattle
University of Florida, Gainesville

Newcastle, Australia **February, 2010**

Building Further in the Stochastic Framework

Probability space: $(\Omega, \mathcal{F}, P)$, elements ω are "future states"

random variables: $X : \Omega \to \boldsymbol{R}$, $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$

typical orientation: $X(\omega)$ = some "cost" or "loss"

Quantification of risk: $\mathcal{R}(X)$ = numerical surrogate for X

$\mathcal{R} : \mathcal{L}^2 \to (-\infty, \infty]$ is then a "risk measure"

Complementary idea: $\mathcal{D}(X)$ = assessment of nonconstancy of X

$\mathcal{D} : \mathcal{L}^2 \to [0, \infty]$ is then a "deviation measure"

standard deviation as a basic example: $\mathcal{D}(X) = \sigma(X)$

Why generalize? motivations in finance, in particular

- **asymmetry** could be beneficial, $\mathcal{D}(-X) \neq \mathcal{D}(X)$?
- promotion of **coherency** in risk (connections will emerge)

Closely related notion: $\mathcal{E}(X)$ = assessment of nonzeroness of X

$\mathcal{E} : \mathcal{L}^2 \to [0, \infty]$ is then an "error measure"

Quantification of Uncertainty

functionals $\mathcal{D} : X \to \mathcal{D}(X) \in [0, \infty]$ for $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$

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)

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

Motivations Coming From Finance

$Y_1, \ldots, Y_m$ = rates of return of various financial instruments
$x_1, \ldots, x_m$ = weights of these instruments in a portfolio
weighting constraints: $(x_1, \ldots, x_m) \in S$ (various versions)
$Y(x_1, \ldots, x_m) = x_1 Y_1 + \cdots + x_m Y_m$ = portfolio rate of return

Classical portfolio problem

Choose the weighting vector $(x_1, \ldots, x_m) \in S$ so as to minimize
$\sigma(\, Y(x_1, \ldots, x_m)\,)$ subject to having $\mu(\, Y(x_1, \ldots, x_m)\,) \geq c$
c = some target level of return, treated parametrically

Issues of contention:

- σ penalizes above-average returns like below-average returns
- the μ constraint may be inappropriately feeble

Innovations to explore: with a switch from gains to losses

- replace $\sigma(Y(x_1, \ldots, x_m))$ by $\mathcal{D}(-Y(x_1, \ldots, x_m))$
- replace $\mu(Y(x_1, \ldots, x_m)) = c$ by $\mathcal{R}(-Y(x_1, \ldots, x_m)) \leq -c$

Estimation Through Linear Regression

Theme: linear approximation of a random variable Y by some other random variables $X_1, \ldots, X_n$ and a constant term

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

"best" coefficients $c_0, c_1, \ldots, c_n$ are to be determined

Existing approaches:

- Classical regression ("least-squares" method)
- Quantile regression (for estimating quantiles/percentiles)
- Modified least squares (Huber approach to outliers)

Issues motivating additional work :

Should "risk preferences" dictate the form of approximation?
Underestimates worse than overestimates for $Y =$ loss/cost?

Quantification of Error in Approximation

Orientation: $X(\omega)$ now 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) $\mathcal{E}(\lambda X) = \lambda\mathcal{E}(X)$ for $\lambda > 0$ (positive homogeneity)

Error measures in the **extended** sense: (E4) dropped

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

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

Some specific instances:

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

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

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

Formulation of Generalized Regression

Let $Y, X_1, \ldots, X_n$ be random variables in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$
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 $c_0, c_1, \ldots, c_n$ in order 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 $(\bar{c}_0, \bar{c}_1, \ldots, \bar{c}_n)$ always exist, and they form a compact convex subset of $\boldsymbol{R}^{n+1}$

Portfolio Motivations Revisited

$Y_1, \ldots, Y_m$ = rates of return of various instruments
$x_1, \ldots, x_m$ = weights of these instruments in a 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 $\hat{Y}_i = c_{i0} + c_{i1}X_1 + \cdots + c_{in}X_n$
$Y(x_1, \ldots, x_m)$ thus replaced in optimization by $\hat{Y}(x_1, \ldots, x_m)$

Serious issue: $(c_{i0}, c_{i1}, \ldots, c_{in})$ can't depend on $(x_1, \ldots, x_m)$!
Should "preferences" therefore influence the mode of regression?

Error Projection

for $\mathcal{E}$ = any measure of error (satisfying the axioms)

THEOREM: deviation measures from error measures

In terms of constants $C \in R$, let

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

Then
- $\mathcal{D}$ is a deviation measure (satisfying the axioms)
- $\mathcal{S}(X)$ is a nonempty closed interval (singleton?)

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

Inverse question: Is every $\mathcal{D}$ the projection of some $\mathcal{E}$?

Yes! but without uniqueness e.g. $\mathcal{E}(X) = \mathcal{D}(X) + |EX|$

Mixture result:

Suppose $\mathcal{D} = \lambda_1 \mathcal{D}_1 + \cdots + \lambda_r \mathcal{D}_r$ with $\mathcal{D}_k$ projected from $\mathcal{E}_k$.

Then $\exists$ "natural" $\mathcal{E}$ built from the $\mathcal{E}_k$'s that projects onto $\mathcal{D}$

but $\mathcal{E} \neq \lambda_1 \mathcal{E}_1 + \cdots + \lambda_r \mathcal{E}_r$

Some Examples of Regression

Classical regression ("least squares")

$$\mathcal{E}(X) = \lambda||X||_2 \text{ for some } \lambda > 0$$
$$\mathcal{S}(X) = \mu(X) = EX \quad \text{mean}$$
$$\mathcal{D}(X) = \lambda\sigma(X) \quad \text{standard deviation, scaled}$$

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

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

Quantiles and Quantile Regression

recall: F_X = c.d.f. for X, $\quad F_X(z) = \text{prob}(X \le z)$

Quantile interval for $\alpha \in (0,1)$: $q_\alpha(X) = [q_\alpha^-(X), q_\alpha^+(X)]$,

$q_\alpha^-(X) = \inf\{x \mid F_X(x) \ge \alpha\}, \quad q_\alpha^+(X) = \sup\{x \mid F_X(x) \le \alpha\}$

Pure quantile regression

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

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

$\mathcal{D}(X) = \text{CVaR}_\alpha(X - EX)$ α-CVaR deviation

Mixed quantile regression (levels α_k, weights $\lambda_k > 0$, $\sum_k \lambda_k = 1$)

$\mathcal{E}(X) = \min\left\{ \sum_{k=1}^r \frac{\lambda_k}{1-\alpha_k} E[X - C_k]^+ - EX \,\middle|\, \sum_{k=1}^r C_k = 0 \right\}$

$\mathcal{S}(X) = \sum_{k=1}^r \lambda_k q_{\alpha_k}(X)$ mixed quantile

$\mathcal{D}(X) = \sum_{k=1}^r \lambda_k \text{CVaR}_{\alpha_k}(X - EX)$ mixed CVaR deviation

Regression Analysis

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

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

$$Z_0(c_1, \ldots, c_n) = Y - [c_1 X_1 + \cdots + c_n X_n] \quad (c_0 \text{ omitted})$$

Regression problem for error measure $\mathcal{E}$:

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

THEOREM: error-shaping decomposition

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

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

$$\bar{c}_0 \in \mathcal{S}(\, Z_0(c_1, \ldots, c_n)\,)$$

COROLLARY: equivalent view 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

Regression error being shaped: through $c_0, c_1, \ldots, c_n$

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

1. **Classical regression**

   minimize $\sigma(Z)$ subject to $\mu(Z) = 0$

2. **Range regression**

   minimize breadth of range of Z subject to the center being 0

3. **Median regression**

   minimize $E|Z|$ subject to "the median of Z being 0"

4. **Quantile regression**

   minimize $\mathcal{D}_\alpha(Z)$ subject to "$q_\alpha(Z) = 0$"

   $$\mathcal{D}_\alpha(Z) = \mathrm{CVaR}_\alpha(Z - EZ)$$

5. **Mixed quantile regression**

   minimize $\sum_k \lambda_k \mathcal{D}_{\alpha_k}(Z)$ subject to "$\sum_k \lambda_k q_{\alpha_k}(Z) = 0$"

Portfolio Application

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

Portfolio rate of return:

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

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: 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: prescription for best $\mathcal{D}$ approximation

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: prescription for best $\mathcal{R}$ approximation

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:

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

New Insights For Regression

- Different approaches to generalized linear regression are deeply connected with different preferences about which approximation error "statistic" should be fixed at 0, and how the deviation from that "statistic" should be shaped
- In a portfolio optimization problem recast in terms of factors, each $\mathcal{D}$ or $\mathcal{R}$ expression naturally suggests its own choice of regression, if the aim is to keep the substitute problem as close as possible to the given problem
- The common practice of generating factor approximations
$$Y_i \approx c_{i0} + c_{i1}X_1 + \cdots + +c_{in}X_n \quad i = 1, \ldots, m,$$
only by "least-squares" regression may lead, when applied in problems of optimization, to risks that are "unacceptable"

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.

THE FUNDAMENTAL QUADRANGLE OF RISK

relating quantifications of various aspects of a random variable

---

risk $\mathcal{R} \longleftrightarrow \mathcal{D}$ deviation

optimization $\quad \uparrow \quad \mathcal{S} \quad \uparrow \quad$ estimation

regret $\mathcal{V} \longleftrightarrow \mathcal{E}$ error

Lecture 1: optimization, the role of $\mathcal{R}$

Lecture 2: estimation, the roles of $\mathcal{E}$, $\mathcal{D}$, $\mathcal{S}$

Lecture 3: tying both together along with $\mathcal{V}$ and duality

Lecture 3

RISK VERSUS DEVIATION, REGRET, AND ENTROPIC DUALITY

R. T. Rockafellar
University of Washington, Seattle
University of Florida, Gainesville

Newcastle, Australia **February, 2010**

Aversity in Risk

toward a fundamental connection with deviation measures

Recall axioms for coherent measures of risk

(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)

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

Another important category of risk 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: (R5) yes, **extended** sense: (R5) no

Risk Measures Paired With Deviation Measures

- Many risk measures are both coherent and averse

  $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X), \quad \mathcal{R}(X) = \sup X$

- Some risk measures are coherent but not averse

  $\mathcal{R}(X) = EX, \quad \mathcal{R}(X) = X(\bar{\omega})$

- Some risk measures are averse but not coherent

  $\mathcal{R}(X) = EX + \lambda\sigma(X)$ (to be seen shortly)

Coherency in deviation: require $\mathcal{D}(X) \leq \sup X - EX$ for all X

THEOREM: deviation versus risk

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

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

where moreover $\mathcal{R}$ is coherent $\iff$ $\mathcal{D}(X)$ is coherent

Note: coherency fails for deviation measures $\mathcal{D}(X) = \lambda\sigma(X)$!

$\Longrightarrow$ risk measures $\mathcal{R}(X) = \mu(X) + \lambda\sigma(X)$ aren't coherent

Safety Margins Revised

Recall the traditional approach to $\mu(X)$ being "safely" below 0:

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

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

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

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

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

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 $\quad Q \in \mathcal{Q} \quad \Longrightarrow \quad Q \geq 0,\; EQ = 1$

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

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

Theorem: basic dualization

$\exists$ **one-to-one** correspondence $\mathcal{R} \longleftrightarrow \mathcal{Q}$ between coherent risk measures $\mathcal{R}$ in the **basic** sense and coherent risk envelopes Q:

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

Conclusion: basic coherency = "customized" worst-case analysis

Some Risk Envelope Examples

recall that "1" $=$ density Q of underlying P with respect to itself

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

$$\mathcal{R}(X) = \sup X \longleftrightarrow \mathcal{Q} = \{\,\text{all } Q \geq 0,\ EQ = 1\,\}$$

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

$$\mathcal{R}(X) = \sum_{k=1}^r \lambda_k \mathcal{R}_k(X) \longleftrightarrow \mathcal{Q} = \{\,\sum_{k=1}^r \lambda_k Q_k \,\big|\, Q_k \in \mathcal{Q}_k\}$$

Dual characterization of aversity:

- $\mathcal{R} \longleftrightarrow \mathcal{Q}$ as before, but $Q \in \mathcal{Q} \quad \not\Rightarrow \quad Q \geq 0$
- must have $1 \in \mathcal{Q}$ "strictly"

Entropic Characterization of Extended Coherency

what happens for coherent $\mathcal{R}$ without positive homogeneity?

Generalized entropy

Call a functional $\mathcal{I}$ on $\mathcal{L}^2$ an entropic distance when

(I1) $\mathcal{I}$ is convex and lower semicontinuous

(I2) $\mathcal{I}(Q) < \infty \implies Q \geq 0,\ EQ = 1$

(I3) $\inf \mathcal{I} = 0 \implies \mathrm{cl}(\mathrm{dom}\,\mathcal{I})$ is a risk envelope $\mathcal{Q}$

Theorem: extended dualization with conjugacy

$\exists$ **one-to-one** correspondence $\mathcal{R} \longleftrightarrow \mathcal{I}$ between coherent risk measures $\mathcal{R}$ in the **extended** sense and entropic distances $\mathcal{I}$:

$$\mathcal{R}(X) = \sup_{Q}\{E[XQ] - \mathcal{I}(Q)\}, \quad \mathcal{I}(Q) = \sup_{X}\{E[XQ] - \mathcal{R}(X)\}$$

Previous correspondence: $\mathcal{I} =$ "indicator" of $\mathcal{Q}$

Aversity: (I3) demands $\mathcal{I}(1) = 0$ with $1 \in \mathcal{Q}$ "strictly"

A Particularly Interesting Example

A pairing with Bolzmann-Shannon entropy

$\mathcal{R}(X) = \log E[e^X]$ coherent and averse corresponds to

$\mathcal{I}(Q) = E[Q \log Q]$ when $Q \geq 0,\ EQ = 1$ but $= \infty$ otherwise

How does this fit into the fundamental quadrangle?

- $\mathcal{D}(X) = \log E[\, e^{(X - EX)}\,]$ deviation measure paired with $\mathcal{R}$
- $\mathcal{E}(X) = E[e^X - X - 1]$ error measure projecting to $\mathcal{D}$
- $\mathcal{S}(X) = \log[e^X] = \mathcal{R}(X)\,!$ the "statistic" associated with $\mathcal{E}$

$\longrightarrow$ some development to be pursued in regression?

Expected Utility

Utility in finance: having a big role in traditional theory

X = incoming money in future, random variable

$u(x)$ = "utility" (in present terms) of getting future amount x

u generally concave, nondecreasing

$u(X(\omega))$ = utility of amount received in state $\omega \in \Omega$

$E[u(X)]$ = expected utility, something to consider maximizing

Importance of a threshold: X = gain/loss against **benchmark**

incrementally, people hate losses more than they love gains!

Normalization of utility: $x > 0$ rel. gain, $x < 0$ rel. loss

$u(0) = 0,\ u'(0) = 1$ for differentiable u, but the latter is equivalent without differentiability to $u(x) \leq x$ for all x

Resulting interpretation:

$u(x)$ = the amount of present money deemed to be acceptable in lieu of getting the future amount x

Translation to Minimization Framework

Utility replaced by regret: $v(x) = -u(-x)$

$v(x)$ = the regret in contemplating a future loss x

= the amount of present money deemed necessary as compensation for a relative loss x in the future

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

Converted context:

X = relative loss in future, random variable

$E[v(X)]$ = expected regret   something to consider minimizing

Insurance interpretation:

$E[v(X)]$ = the amount to charge (with respect to v) for covering the uncertain future loss X

Observations: about $\mathcal{V}(X) = E[v(X)]$ as a functional on $\mathcal{L}^2$

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

Quantifications of Regret in General

expressions $\mathcal{V}(X)$ for potential losses X, not just of form $E[v(X)]$

Coherency in regret

Call $\mathcal{V}$ a **coherent** measure of regret if

(V1) $\mathcal{V}(0) = 0$

(V2) $\mathcal{V}((1-\lambda)X + \lambda X') \leq (1-\lambda)\mathcal{V}(X) + \lambda\mathcal{V}(X')$ (convexity)

(V3) $\mathcal{V}(X) \leq \mathcal{V}(X')$ when $X \leq X'$ (monotonicity)

(V4) $\mathcal{V}(X) \leq c$ when $X_k \to X$ with $\mathcal{V}(X_k) \leq c$ (closedness)

(V5) $\mathcal{V}(\lambda X) = \lambda\mathcal{V}(X)$ for $\lambda > 0$ (positive homogeneity)

Aversity in regret

Call $\mathcal{V}$ an **averse** measure of regret if (V3) is relinquished, but

(V6) $\mathcal{V}(X) > EX$ for all nonconstant X (aversity)

basic sense: (V5) yes, **extended** sense: (V5) no

A Trade-off That Identifies Risk

For $\mathcal{V}$ = some measure of regret consider the expression:

$C + \mathcal{V}(X - C)$ for a future loss X and constants C

Interpretation: accept a certain loss C, thereby shifting the threshold and only regetting a residual future loss $X - C$

Theorem: derivation of risk from regret

Given an **averse** regret measure $\mathcal{V}$, define $\mathcal{R}$ and $\mathcal{S}$ by

$$\mathcal{R}(X) = \min_C \{C + \mathcal{V}(X - C)\}, \quad \mathcal{S}(X) = \operatorname{argmin}_C \{C + \mathcal{V}(X - C)\}$$

Then
- $\mathcal{R}$ is an **averse** risk measure (coherent for $\mathcal{V}$ coherent)
- $\mathcal{S}(X)$ is a nonempty closed interval (singleton?)

CVaR example: $\mathcal{V}(X) = E[\frac{1}{1-\alpha} X_+]$

$$\mathcal{R}(X) = \min_C \{C + \tfrac{1}{1-\alpha} E[X - C]_+\} = \mathrm{CVaR}_\alpha(X)$$

$\longrightarrow$ the key minimization rule with argmin $= \mathrm{VaR}_\alpha(X) = q_\alpha(X)$

Completing the Fundamental Quadrangle of Risk

Error versus regret

The simple relations

$$\mathcal{E}(X) = \mathcal{V}(X) - EX, \quad \mathcal{V}(X) = EX + \mathcal{E}(X),$$

provide a **one-to-one** correspondence between error measures $\mathcal{E}$ and **averse** regret measures $\mathcal{V}$ (with $V(C) < \infty$?), where

$$\mathcal{V} \text{ is coherent} \iff \mathcal{E}(-X) \leq EX \text{ when } X \geq 0$$

Moreover, the $\mathcal{R}$ from $\mathcal{V}$ is **paired** with the $\mathcal{D}$ from $\mathcal{E}$, and in the minimization formulas giving statistics $\mathcal{S}$,

$$\text{the } \mathcal{S}(X) \text{ from } \mathcal{V} \to \mathcal{R} \quad = \quad \text{the } \mathcal{S}(X) \text{ from } \mathcal{E} \to \mathcal{D}$$

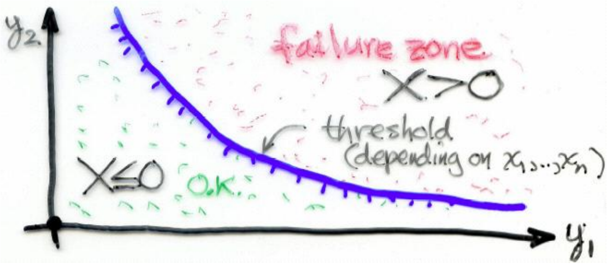
Expectation version:

$$\mathcal{V}(X) = E[v(X)] \longleftrightarrow \mathcal{E}(X) = E[\varepsilon(X)]$$
$$\varepsilon(x) = v(x) - x, \quad v(x) = x + \varepsilon(x)$$

Further Development From an Engineering Perspective

Uncertain "cost": $X = c(x_1, \ldots, x_n; Y_1, \cdots, Y_r)$

$x_1, \ldots, x_n$ =design variables, $Y_1, \ldots, Y_r$ =stochastic parameters



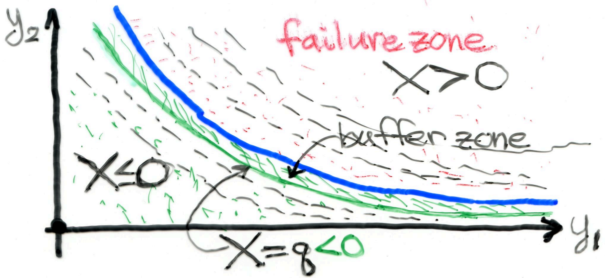
Probability of failure: $p_f = \text{prob}\{X > 0\}$

- How to compute or at least estimate?
- How to cope with dependence on $x_1, \ldots, x_n$ in optimization?

  Both p_f and the threshold **shift** with changes in $x_1, \ldots, x_n$

Buffered Failure — Enhanced Safety

Uncertain "cost": $X = c(x_1, \ldots, x_n; Y_1, \cdots, Y_r)$



Buffered probability of failure: $P_f = \text{prob}\,\{X > q\}$

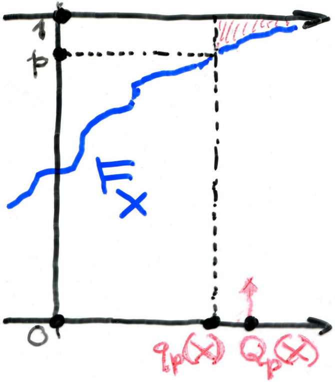
q determined so as to make $E[\,X \,|\, X > q\,] = 0$

Suggestion: adjust failure modeling to P_f in place of p_f

safer by integrating tail information, and

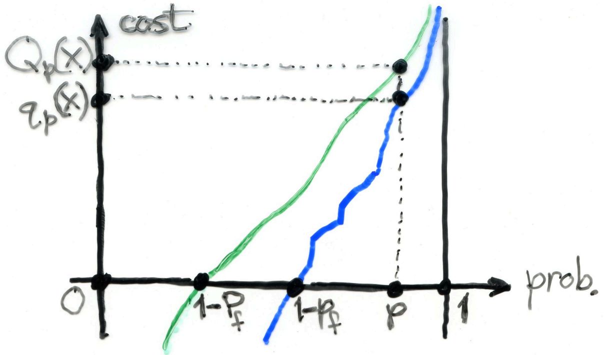
easier also to work with in optimization!

Quantiles and "Superquantiles"



quantile: $q_p(X) = F_X^{-1}(p) = \mathrm{VaR}_p(X)$
superquantile: $Q_p(X) = E[\,X\,|X > q_p(X)\,] = \mathrm{CVaR}_p(X)$
terms in finance: value-at-risk and conditional value-at-risk

Diagram of Relationships



$$q_p(X) = F_X^{-1}(p), \qquad Q_p(X) = \frac{1}{1-p}\int_p^1 q_s(X)\,ds$$

$q_p(X)$ can depend poorly on p, but $Q_p(X)$ depends smoothly on p

failure modeling: p_f determined by $q_p(X) = 0,\ p = 1 - p_f$

P_f determined by $Q_p(X) = 0,\ p = 1 - P_f$

Comparison of Roles in Optimization

Key fact: $\mathcal{R}(X) = Q_p(X)$ is **coherent** but $\mathcal{R}(X) = q_p(X)$ is not!

Constraint $p_f(\, c(x_1, \ldots, x_n, Y_1, \ldots, Y_m)\,) \leq 1 - p$ corresponds to
$q_p(\, c(x_1, \ldots, x_n, Y_1, \ldots, Y_m)\,) \leq 0$

Constraint $P_f(\, c(x_1, \ldots, x_n, Y_1, \ldots, Y_m)\,) \leq 1 - p$ corresponds to
$Q_p(\, c(x_1, \ldots, x_n, Y_1, \ldots, Y_m)\,) \leq 0$

Minimizing $q_p(\, c(x_1, \ldots, x_n, Y_1, \ldots, Y_m)\,)$ corresponds to
finding $x_1, \ldots, x_n$ with lowest C such that
$c(x_1, \ldots, x_n, Y_1, \ldots, Y_m) \leq C$ with probability $< 1 - p$

Minimizing $Q_p(\, c(x_1, \ldots, x_n, Y_1, \ldots, Y_m)\,)$ corresponds to
finding $x_1, \ldots, x_n$ with lowest C such that, even in the $1 - p$
worst fraction of cases, $c(x_1, \ldots, x_n, Y_1, \ldots, Y_m) \leq C$ on average

Some References

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

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

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

[4] R. T. Rockafellar, J. O. Royset (2010), "On buffered failure probability in design and optimization of structures, *Journal of Reliability Engineering and System Safety.* downloadable:
faculty.nps.edu/joroset/docs/RockafellarRoyset_RESS.pdf