Optimization and Statistical Estimation

support for decision-making in a stochastic environment

Optimization:
minimize a “cost” expression under constraints on the decision
the constraints could involve bounds on other “costs”
the “costs” may have a background in statistical analysis

Statistical Estimation:
approximate some quantity from empirical/historical data
minimize an error expression to get regression coefficients
different interpretations of “error” yield different results

Interplay:
- optimization problems involving uncertainty depend on
  estimation methodology even in coming to a formulation
- estimation problems are optimization of a special sort
- new and deeper connections are now coming to light
The Risk Quadrangle — A New Paradigm

an array of “quantifications” be applied to random variables \( X \)
“cost” orientation of \( X \): high outcomes bad, low outcomes good

\[
\begin{align*}
\text{risk} & \quad \mathcal{R} \quad \leftrightarrow \quad \mathcal{D} \quad \text{deviation} \\
\text{optimization} & \quad \uparrow \downarrow \quad \mathcal{S} \quad \downarrow \uparrow \quad \text{statistics} \\
\text{regret} & \quad \mathcal{V} \quad \leftrightarrow \quad \mathcal{E} \quad \text{error}
\end{align*}
\]

\( \mathcal{R}(X) \) elicits the “level of cost” in \( X \) to use in making comparisons
\( \mathcal{V}(X) \) quantifies the “anti-utility” in outcomes \( X > 0 \) versus \( X \leq 0 \)
\( \mathcal{D}(X) \) measures the “nonconstancy” in \( X \) as its uncertainty
\( \mathcal{E}(X) \) measures the “nonzeroness” in \( X \) for use in estimates
\( \mathcal{S}(X) \) is a “statistic” associated with \( X \) through \( \mathcal{E} \) and through \( \mathcal{V} \)
R.T. Rockafellar and S.P. Uryasev (2013),
“The fundamental risk quadrangle in risk management, optimization and statistical estimation,”

downloads: www.math.washington.edu/~rtr/mypage.html
(look for item #218)
“Costs”: quantities to be minimized or kept below given levels (this fits better with optimization conventions than “profits”)

General “cost” expression in decision-making:
\[ c(x, v) \text{ with } x = \text{decision vector}, \ v = \text{data vector} \]
\[ x = (x_1, \ldots, x_n), \quad v = (v_1, \ldots, v_m) \]

Stochastic uncertainty:
\[ v \text{ is replaced by a random variable vector } V = (V_1, \ldots, V_m) \]
then “cost” becomes a random variable:
\[ c(x) = c(x, V) \]

Portfolio example in finance:
\[ V_j = \text{random return on asset } j, \ x_j = \text{amount of } j \text{ in portfolio} \]
\[ c(x) = c(x, V) = -[x_1 V_1 + \cdots + x_n V_n] \text{ (random loss incurred)} \]

Design examples in engineering: “cost” $\leftrightarrow$ “hazard”
Challenges in Optimization Modeling

“cost” \( c(x) \) = random variable depending on decision \( x \)
outcomes \( < 0 \), if any, correspond to “rewards”

Key issue in problem formulation

- the distribution of \( c(x) \) can only be shaped by the choice of \( x \)
- but how then can constraints and minimization be understood?
A Broad Pattern for Handling Risk in Optimization

Risk measures: functionals $\mathcal{R}$ that “quantify the risk” in a random variable $X$ by a numerical value $\mathcal{R}(X)$ (“risk” $\neq$ “uncertainty”)

Systematic prescription

Faced with an uncertain “cost” $c(x) = c(x, V)$ articulate it numerically as $\bar{c}(x) = \mathcal{R}(c(x))$ for a choice of risk measure $\mathcal{R}$

Constraints: keeping $c(x)$ “acceptably” $\leq b$
  modeled as: constraint $\bar{c}(x) = \mathcal{R}(c(x)) \leq b$

Objectives: making $c(x)$ as “acceptably” low as possible
  modeled as: minimizing $\bar{c}(x) = \mathcal{R}(c(x))$
  i.e., minimizing the threshold level $b$ such that $x$ can be selected with $c(x)$ “acceptably” $\leq b$
Stochastic Framework

Space of future states: \( \Omega \) with elements \( \omega \) ("scenarios")
\[ A_0 = \text{field of subsets of } \Omega, \quad P_0 = \text{probability measure on } A_0 \]

Random variables: functions \( X : \Omega \to \mathbb{R} \) (\( A_0 \)-measurable)
\[ EX = \mu(X) = \int_{\Omega} X(\omega) dP_0(\omega), \quad \sigma^2(X) = E[(X - EX)^2] \]

Function space setting: \( X \in L^2 := L^2(\Omega, A_0, P_0) \) for simplicity
Hilbert space of random variables with finite mean and variance
\[ \langle X, Y \rangle = E[XY], \quad \|X\| = \sqrt{E[X^2]} \]

Treatment of probability alternatives

Measures \( P \) can be represented by densities \( \frac{dP}{dP_0} \) with respect to \( P_0 \)
\[ E_P(X) = \int_{\Omega} X(\omega) \frac{dP}{dP_0}(\omega) dP_0(\omega) = \langle X, Q \rangle \quad \text{for } Q = \frac{dP}{dP_0} \]
[sets \( P \) of alternatives \( P \) \( \longleftrightarrow \) [sets \( Q \) of \( Q \in L^2: Q \geq 0, EQ = 1 \)]
Axiomatization of Risk

**Axioms for regular measures of risk:** \( \mathcal{R} : \mathcal{L}^2 \to (-\infty, \infty] \)

- (R1) \( \mathcal{R}(C) = C \) for all constants \( C \)
- (R2) convexity, (R3) closedness (lower semicontinuity)
- (R4) aversity: \( \mathcal{R}(X) > EX \) for nonconstant \( X \)

**Additional properties of major interest:**

- (R5) positive homogeneity: \( \mathcal{R}(\lambda X) = \lambda \mathcal{R}(X) \) when \( \lambda > 0 \)
- (R6) monotonicity: \( \mathcal{R}(X) \leq \mathcal{R}(X') \) when \( X \leq X' \)

**Coherent measures of risk:** \( \mathcal{R} \) satisfying (R1), (R2), (R5), (R6)

Artzner et al. (2000) introduced coherency without aversity

**Note:** (R1)+(R2) \( \implies \mathcal{R}(X + C) = \mathcal{R}(X) + C \) for constants \( C \)

**Preservation of convexity under (R1)+(R2)+(R6)**

\( c(x) = c(x, V) \) convex in \( x \) \( \implies \bar{c}(x) = \mathcal{R}(c(x)) \) convex in \( x \)
Some Common Approaches From This Perspective

Best guess of future state: \( R(X) = X(\bar{\omega}) \) (\( \text{prob}(\bar{\omega}) > 0 \))

\[
\begin{align*}
\text{then } \ R(X) \leq b & \iff X(\bar{\omega}) \leq b \\
\rightarrow \text{coherent but not averse (and lacking any ability to hedge)}
\end{align*}
\]

Focusing on worst cases: \( R(X) = \sup X \) (ess. sup)

\[
\begin{align*}
\text{then } \ R(X) \leq b & \iff X \leq b \text{ almost surely} \\
\rightarrow \text{coherent, averse (but perhaps overly conservative, infeasible)}
\end{align*}
\]

Passing to expectations: \( R(X) = \mu(X) = \mathbb{E}X \)

\[
\begin{align*}
\text{then } \ R(X) \leq b & \iff X \leq b \text{ “on average”} \\
\rightarrow \text{coherent but not averse (perhaps too feeble)}
\end{align*}
\]

Adopting a safety margin: \( R(X) = \mu(X) + \lambda \sigma(X) \) \( \lambda > 0 \)

\[
\begin{align*}
\text{then } \ R(X) \leq b & \text{ unless in tail } > \lambda \text{ standard deviations} \\
\rightarrow \text{averse but not coherent (lacks monotonicity!)}
\end{align*}
\]
Quantiles and “Superquantiles”: VaR and CVaR

\[ F_X = \text{cumulative distribution function for random variable } X \]

**p-Quantile:** “value-at-risk” in finance

\[ q_p(X) = \text{VaR}_p(X) = “F_X^{-1}(p)” \]

**p-Superquantile:** “conditional value-at-risk” in finance

\[ Q_p(X) = \text{CVaR}_p(X) = “E\[X | X \geq q_p(X)\]” = \frac{1}{1-p} \int_p^1 q_t(X)dt \]

**mathematical behavior:** quantiles bad, superquantiles good
Measures of Risk Based on Probability Thresholds

Looking at quantiles/VaR: \( R(X) = q_p(X) \)
then \( R(X) \leq b \iff \text{prob}\{X \leq b\} \geq p \)
\( \rightarrow \) not coherent, not averse (troublesome, subject to criticism)

Superquantiles/CVaR instead: \( R(X) = Q_p(X) \)
then \( R(X) \leq b \iff c(x) \leq b \) on average in upper \( p \)-tail
\( \rightarrow \) coherent, averse (easy to work with and more conservative!)

Corresponding concepts of “failure”
\( c(x) = c(x, V) \) captures “hazard,” failure \( \iff \) outcomes \( > 0 \)
\( q_p(c(x)) \leq 0 \) means **ordinary** probability of failure is \( \leq 1 - p \)
\( Q_p(c(x)) \leq 0 \) means **buffered** probability of failure is \( \leq 1 - p \)

Example: case of \( p = 0.9 \), focusing on the worst 10% of events
\( \rightarrow \) buffered probability of failure \( \leq 0.1 \) means: even in that
tail range, the hazard variable comes out “safe on average”
Minimization Formula for VaR and CVaR

\[
\text{CVaR}_p(X) = \min_{C \in \mathbb{R}} \left\{ C + \frac{1}{1-p} E[\max\{0, X - C\}] \right\} \quad \text{for } p \in (0, 1)
\]

\[
\text{VaR}_p(X) = \arg\min \quad \text{(if unique, otherwise the lowest)}
\]

Application to CVaR models:

convert a problem in \( x \) like

\[
\text{minimize } \text{CVaR}_{p_0}(c_0(x)) \quad \text{subject to} \quad [\text{basic constraints and}]
\]

\[
\text{CVaR}_{p_i}(c_i(x)) \leq b_i \quad \text{for } i = 1, \ldots, m
\]

into a problem in \( x \) and auxiliary variables \( C_0, C_1, \ldots, C_m \),

\[
\text{minimize } C_0 + \frac{1}{1-p_0} E[\max\{0, c_0(x) - C_0\}] \quad \text{while requiring}
\]

\[
C_i + \frac{1}{1-p_i} E[\max\{0, c_i(x) - C_i\}] \leq b_i, \quad i = 1, \ldots, m
\]

Important case: this converts to linear programming when

(1) each \( c_i(x) = c_i(x, V) \) depends linearly on \( x \),

(2) the future state space \( \Omega \) is modeled as finite.
Some CVaR/Superquantile References


downloads: www.math.washington.edu/~rtr/mypage.html
look for items #179, #187, #211
Stochastic Ambiguity, Admitting Alternative Probabilities

Probability density functions: \( Q \in \mathcal{L}^2 \) with \( Q \geq 0, \) \( EQ = 1 \)

\[ \implies Q = \frac{dP}{dP_0} \] for some probability measure \( P \)

\[ E_P(X) = \langle X, Q \rangle = E[XQ] = \int_\Omega X(\omega) \frac{dP}{dP_0}(\omega) dP_0(\omega) \]

the underlying probability measure \( P_0 \) corresponds to \( Q \equiv 1 \)

Stochastic ambiguity: not trusting just \( P_0 \), looking at other \( P \)
interested in \( \sup_{P \in \mathcal{P}} E_P(X) \) instead of just \( EX = E_{P_0}(X) \)

Risk envelopes: sets \( Q \subset \mathcal{L}^2 \) consisting of probability densities \( Q \)
interested in \( \mathcal{R}(X) = \sup_{Q \in Q} E[XQ] \) as a measure of risk

note: \( \mathcal{R} \) is unaffected if \( Q \) replaced by its closed convex hull

Regularity of a risk envelope:
\( Q \) is closed convex \( \neq \emptyset \) and \( 1 \in Q \), but \( 1 \) isn’t a “support point”
(i.e., \( \not\exists \) hyperplane touching \( Q \) at \( Q \equiv 1 \) without \( \supset Q \))
Dualization of Monotonic Measures of Risk

Risk envelope characterization, positively homogeneous case

\[ \mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} E[XQ], \quad \mathcal{Q} = \{ Q \mid E[XQ] \leq \mathcal{R}(X), \forall X \} \]

furnishes a one-to-one correspondence between

(a) regular risk measures \( \mathcal{R} \) that are monotonic + pos. homog.
(b) regular risk envelopes \( Q \), as above

Risk envelope characterization, general case

\[ \mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} \{ E[XQ] - \mathcal{J}(Q) \}, \]
\[ \mathcal{J}(Q) = \sup_{X \in L^2} \{ E[XQ] - \mathcal{R}(Q) \}, \quad Q = \text{cl}(\text{dom } \mathcal{J}) \]

furnishes a one-to-one correspondence between

(a) regular risk measures \( \mathcal{R} \) that are monotonic
(b) regular risk envelopes \( Q \) as \( \text{cl}(\text{dom } \mathcal{J}) \) for a closed convex functional \( \mathcal{J} : L^2 \to [0, \infty] \) such that \( \mathcal{J}(1) = 0 = \min \mathcal{J} \)

\( \mathcal{J}(Q) \) assesses the “divergence” of \( Q = \frac{dP}{dP_0} \) from \( 1 = \frac{dP_0}{dP_0} \)
Some Examples of Risk Dualization

Risk measures with positive homogeneity:

- for $\mathcal{R}(X) = Q_p(X) = \text{CVaR}_p(X)$ the risk envelope is:
  
  $Q = \{ Q \mid Q \geq 0, \, EQ = 1, \, Q \leq \frac{1}{1-p} \}$

- for $\mathcal{R}(X) = \sup X$ the risk envelope is:
  
  $Q = \{ Q \mid Q \geq 0, \, EQ = 1 \}$ (the full “probability simplex”)

Risk measures without positive homogeneity:

$\mathcal{R}(X) = \log E[\exp X]$ is a regular measure of risk that is also monotonic but not positively homogeneous. Its risk envelope is $Q = \text{probability simplex}$ supplied with $\mathcal{J}(Q) = -E[Q \log Q]$

Note: for $Q = dP/P_0$, the Bolzano-Shannon entropy expression

$$-E[Q \log Q] = -\int_{\Omega} \left[ \frac{dP}{dP_0}(\omega) \log \frac{dP}{dP_0}(\omega) \right] dP_0(\omega)$$

is known as the Kullback-Leibler divergence of $P$ from $P_0$. 
“Robust” Optimization Revisited and Refined

Motivation behind so-called “robust” optimization:
- probabilities are often hard to assess, even as guesswork
- they can be avoided by focusing on $\mathcal{R}(X) = \sup_{\omega \in \Omega} X(\omega)$

Practical compromise/pitfall: this depends on the model for $\Omega$
- which “scenarios” $\omega$ should go into $\Omega$ as the uncertainty set?
- subjective probability enters in deciding which ones to exclude

Nested robustness

Let $\Omega$ be **partitioned** into $\Omega_1, \ldots, \Omega_N$ with $p_k = \text{prob}[\Omega_k]$, and let

$$\mathcal{R}(X) = \sum_{k=1}^{N} p_k \sup_{\omega \in \Omega_k} X(\omega).$$

Then $\mathcal{R}$ is regular, monotonic, pos. homogeneous, with envelope

$$Q = \{ Q \mid Q \geq 0, \text{prob}_Q[\Omega_k] = p_k, \forall k \}.$$ 

Alternative interpretation:
- the partition generates an information field $\mathcal{A}$
- $Q \leftrightarrow$ all prob. measures $P$ consistent with that information
Recalling the Risk Quadrangle for the Next Development

\[
\begin{align*}
\text{risk } & \mathcal{R} \leftrightarrow \mathcal{D} \quad \text{deviation} \\
\text{optimization} & \uparrow \downarrow \quad \mathcal{S} \quad \downarrow \uparrow \\
\text{regret } & \mathcal{V} \leftrightarrow \mathcal{E} \quad \text{error}
\end{align*}
\]

\(\mathcal{R}(X)\) elicits the “level of cost” in \(X\) to use in making comparisons
\(\mathcal{V}(X)\) quantifies the “anti-utility” in outcomes \(X > 0\) versus \(X \leq 0\)
\(\mathcal{D}(X)\) measures the “nonconstancy” in \(X\) as its uncertainty
\(\mathcal{E}(X)\) measures the “nonzeroness” in \(X\) for use in estimates
\(\mathcal{S}(X)\) is a “statistic” associated with \(X\) through \(\mathcal{E}\) and through \(\mathcal{V}\)
Regret Versus Utility

**Regret:** the “compensation” $\mathcal{V}(X)$ for facing a future cost/loss $X$ in contrast to the “utility” $\mathcal{U}(Y)$ perceived in a future gain $Y$

$$\mathcal{V}(X) = -\mathcal{U}(-X) \iff \mathcal{U}(Y) = -\mathcal{V}(-Y)$$

**Axioms for regular measures of regret:** $\mathcal{V} : L^2 \rightarrow (-\infty, \infty)$

- (V1) $\mathcal{V}(0) = 0$, (V2) convexity, (V3) closedness
- (V4) aversity: $\mathcal{V}(X) > EX$ for nonconstant $X$

**Additional properties of major interest:**

- (V5) positive homogeneity: $\mathcal{V}(\lambda X) = \lambda \mathcal{V}(X)$ when $\lambda > 0$
- (V6) monotonicity: $\mathcal{V}(X) \leq \mathcal{V}(X')$ when $X \leq X'$

$\implies$ axioms for **regular measures of utility** $\mathcal{U} : L^2 \rightarrow [-\infty, \infty)$

(more explanation of utility connections will come later)
regret is oriented to minimizing, utility is oriented to maximizing
Risk From Regret

**Goal:** generalize to other risk measures the superquantile formula

\[ Q_p(X) = \min_{C \in \mathbb{R}} \left\{ C + \frac{1}{1-p} E[\max\{0, X - C\}] \right\}, \quad q_p(X) = \arg\min \]

**Trade-off Theorem**

For any regular measure of regret \( \mathcal{V} \), the formula

\[ \mathcal{R}(X) = \min_{C \in \mathbb{R}} \left\{ C + \mathcal{V}(X - C) \right\} \]

defines a regular measure of risk \( \mathcal{R} \) such that

- \( \mathcal{V} \) monotonic \( \implies \mathcal{R} \) monotonic
- \( \mathcal{V} \) pos. homog. \( \implies \mathcal{R} \) pos. homog.

**Trade-off interpretation:**

- \( C = \) “designated loss” (to be written off here and now)
- \( X - C = \) “residual loss” (still to be faced in the future)

**Application to insurance:** the argmin leads to the “premium”

**Optimization role:** simplifying \( \mathcal{R} \) to \( \mathcal{V} \) in objective/constraints
More About Utility

Finance question: for random variables $Y$ representing monetary gains, how to think of one as being preferable to another?

Traditional approach through expected utility

- there is a utility function $u$ to apply to money amounts $y$
- the functional $\mathcal{U} : Y \rightarrow \mathcal{U}(Y) = E[u(Y)]$ is then the key:

$$Y_1 \text{ is preferred (strictly) to } Y_2 \iff \mathcal{U}(Y_1) > \mathcal{U}(Y_2)$$

Background: von Neumann/Morgenstern theory for “lotteries”

- the utility function $u$ is **concave and nondecreasing**, and can be **normalized** to have $u(0) = 0$ and $u(y) \leq y$

$$\implies \text{a benchmark focus with utility scaled to money, in which } \mathcal{U} \text{ is concave, nondecreasing, with } \mathcal{U}(0) = 0 \text{ and } \mathcal{U}(Y) \leq E[Y]$$

Regular measures of utility: such $\mathcal{U}(Y)$ more generally
Utility and Regret in the Monotonic Expectational Case

**Expected utility:** \( U(Y) = E[u(Y)] \) (normalized \( u \))
\( u(y) \) concave, nondecreasing with \( u(0) = 0, u(y) \leq y \)

**Expected regret:** \( V(X) = E[v(X)] \)
\( v(x) \) convex, nondecreasing with \( v(0) = 0, v(x) \geq x \)

\[ v(x) = -u(-x) \quad \leftrightarrow \quad u(y) = -v(-y) \]

**Superquantile formula example:** \( V(X) = \frac{1}{1-p} E[\max\{0, X\}] \)
\[ v(x) = \frac{1}{1-p} \max\{0, x\}, \quad u(y) = \frac{1}{1-p} \min\{0, y\} \]
Measures of Utility Beyond Simple Expected Utility

Utility reflecting stochastic ambiguity

- Let $u$ be a nondecreasing concave utility function, normalized
- For a risk envelope $Q_0$ and an associated divergence $J_0$, let
  \[ U(Y) = \inf_{Q \in Q_0} \{ E[u(Y)Q] + J_0(Q) \} \]
  Then $U$ is a regular measure of utility that is monotonic:
    $U$ is concave, nondecreasing, with $U(0) = 0$ and $U(Y) \leq E[Y]$

Corresponding ambiguity version of regret

- Let $v$ be a nondecreasing convex regret function
- For a risk envelope $Q_0$ and an associated divergence $J_0$, let
  \[ V(X) = \sup_{Q \in Q_0} \{ E[v(X)Q] - J_0(Q) \} \]
  Then $V$ is a regular measure of regret that is monotonic

Relation to risk: for the risk measure $R_0$ dual to $Q_0$, $J_0$,
  \[ U(Y) = -R_0(-u(Y)), \quad V(X) = R_0(v(X)) \]
Passing Now to the Statistics Side of the Risk Quadrangle

\[ risk \; R \leftrightarrow D \; deviation \]

optimization \[ \uparrow \downarrow \; S \; \downarrow \uparrow \] statistics

regret \[ V \leftrightarrow E \; error \]

\( R(X) \) elicits the “level of cost” in \( X \) to use in making comparisons
\( V(X) \) quantifies the “anti-utility” in outcomes \( X > 0 \) versus \( X \leq 0 \)
\( D(X) \) measures the “nonconstancy” in \( X \) as its uncertainty
\( E(X) \) measures the “nonzeroness” in \( X \) for use in estimates
\( S(X) \) is a “statistic” associated with \( X \) through \( E \) and through \( V \)
Deviation as a Quantification of Uncertainty

\( D(X) \) generalizes standard deviation \( \sigma(X) \)

Axioms for regular measures of deviation: \( D : L^2 \rightarrow [0, \infty] \)

(D1) \( D(C) = 0 \) for constant random variables \( C \)
(D2) convexity,  (D3) closedness
(D4) robustness: \( D(X) > 0 \) for nonconstant \( X \)

**Additional properties of major interest:**

(D5) positive homogeneity: \( D(\lambda X) = \lambda D(X) \) when \( \lambda > 0 \)
(D6) upper range boundedness: \( D(X) \leq \sup X - EX \)

**Note:** \((D1)+(D2) \implies D(X + C) = D(X)\) for constants \( C \)
symmetry not assumed, perhaps \( D(-X) \neq D(X) \)

**Example:** \( D(X) = \sigma(X) = ||X - EX|| \) fails to satisfy (D6), but \( D(X) = \sigma_+(X) = ||\max\{0, X - EX\}|| \) satisfies all

**Extended CAPM:** obtained with \( D(X) \) replacing \( \sigma(X) \) (finance)
Risk Versus Deviation

quantification of “cost/loss” versus quantification of uncertainty

Mean±deviation representation of risk measures

A one-to-one correspondence $D \leftrightarrow R$ between regular risk measures $R$ and regular deviation measures $D$ is given by

$$R(X) = EX + D(X), \quad D(X) = R(X - EX),$$

where moreover monotonicity (R6) of risk is characterized by

$R(X)$ satisfies (R6) $\iff D(X)$ satisfies (D6)

Example 1: the risk measure $R(X) = EX + \lambda \sigma(X), \lambda > 0,$ is regular but not monotonic because $D(X) = \lambda \sigma(X)$ fails (D6)

Example 2: the deviation measure $D(X) = CVaR_p(X - EX)$ is not only regular but also, in addition, satisfies (D6)
Deviation From Error

looking now at a concept of “error” that can be asymmetric

**Axioms for regular measures of error:** \( \mathcal{E} : \mathcal{L}^2 \rightarrow [0, \infty] \)

- (E1) \( \mathcal{E}(0) = 0 \),
- (E2) convexity,
- (E3) closedness
- (E4) robustness: \( \mathcal{E}(X) > 0 \) for nonzero \( X \)

**Additional properties of major interest:**

- (E5) positive homogeneity: \( \mathcal{E}(\lambda X) = \lambda \mathcal{E}(X) \) when \( \lambda > 0 \)
- (E6) \( \mathcal{E}(X) \leq |EX| \) when \( X \leq 0 \)

**Error projection (with respect to constants \( C \))**

For a regular error measure \( \mathcal{E} \), let

\[
\mathcal{D}(X) = \min_C \mathcal{E}(X - C), \quad S(X) = \arg\min_C \mathcal{E}(X - C).
\]

Then \( \mathcal{D} \) is a regular deviation measure, \( S \) the associated “statistic”

\( \mathcal{E}(X) \) satisfies (E6) \( \implies \) \( \mathcal{D}(X) \) satisfies (D6)

\( \implies S(X) \) is the constant \( C \) “nearest” to \( X \) with respect to \( \mathcal{E} \)
Some Error/Statistic Examples

Example 1: \( \mathcal{E}(X) = \sqrt{E[X^2]} \) yields \( S(X) = EX \)
this regular measure of error fails to satisfy (E6)

Example 2: \( \mathcal{E}(X) = E|X| \) yields \( S(X) = \text{median of } X \)
this regular measure of error satisfies all

Example 3: \( \mathcal{E}(X) = \sup|X| \) yields \( S(X) = \frac{1}{2}[\sup X + \inf X] \)
this regular measure of error fails to satisfy (E6)

Example 4: \( \mathcal{E}(X) = \frac{1}{1-p} E[\max\{0, X\}] - EX \) yields
the \( p \)-quantile statistic \( S(X) = q_p(X) \)
this asymmetric regular measure of error satisfies all

Example 5: \( \mathcal{E}(X) = E[\exp X - X - 1] \) yields \( S(X) = \log E[\exp X] \)
this asymmetric regular measure of error satisfies all
Available information: e.g. a large collection of pairs \((x_k, y_k)\)
Perspective: empirical distribution in \(x, y\) space of r.v.'s \(X, Y\)
Approximation: \(Y \approx aX + b\), error gap \(Z(a, b) = Y - [aX + b]\)
Regression From a Broader Point of View

\[ Y = \text{random variable (scalar) to be understood in terms of} \]
\[ X_1, \ldots, X_n = \text{some “more basic” variables (e.g., “factors”)} \]

**Approximation scheme:**
\[ Y \approx f(X_1, \ldots, X_n) \quad \text{for} \quad f \in \mathcal{F} \]
\[ \mathcal{F} = \text{some specified class of functions} \quad f : \mathbb{R}^n \rightarrow \mathbb{R} \]
\[ \text{e.g. linear, } f(x_1, \ldots, x_n) = c_0 + c_1 x_1 + \cdots + c_n x_n \]

**Error gap variable:**
\[ Z_f = Y - f(X_1, \ldots, X_n) \quad \text{for} \quad f \in \mathcal{F} \]

**Regression problem, in general**

minimize \( \mathcal{E}(Z_f) \) over all \( f \in \mathcal{F} \) for some error measure \( \mathcal{E} \)

**Standard regression:**
\[ \mathcal{E}(Z_f) = (E[Z_f^2])^{1/2} \quad \text{“least squares”} \]

**Quantile regression:**
using \( Z^+ = \max\{0, Z\}, \quad Z^- = \max\{0, -Z\} \)
\[ \mathcal{E}(Z_f) = E[\frac{p}{1-p}Z_f^+ + Z_f^-] \quad \text{at probability level } p \in (0, 1) \]
Koenker-Bassett error, normalized
Effect of the Choice of Error Measure

error gap variable to be “made small”: \( Z_f = Y - f(X_1, \ldots, X_n) \)

Regression problem “decomposition” (when \( f \in \mathcal{F} \Rightarrow f + C \in \mathcal{F} \))

- minimizing \( \mathcal{E}(Z_f) \) over all \( f \in \mathcal{F} \) corresponds to
- minimizing \( \mathcal{D}(Z_f) \) under the constraint \( S(Z_f) = 0 \)

Important issue for connecting with optimization:

- parameterized “costs” \( c(x) = c(x, V) \) for \( x = (x_1, \ldots, x_n) \)
- can be viewed as \( Y = c(X, V) \) with \( X = (X_1, \ldots, X_n) \)
- \( X = \) “randomized decision” tied to empirical sample at hand

- regression can serve then to get a “formula” for \( c(x) \approx f(x) \)
- for using \( \bar{c}(x) = \mathcal{R}(c(x)) \), shouldn’t this be “tuned” to \( \mathcal{R} \)?
- \( c(X, V) \) may only be supported by some \( (X, V) \) database!
Example: an Application to Composition of Alloys

Alloy model: a mixture of various metals
amounts of chief ingredients: \( x = (x_1, \ldots, x_n) \) “design”
amounts of other ingredients: \( v = (v_1, \ldots, v_m) \) “contaminants”
a “characteristic” to be controlled: \( y \) ideally kept \( \leq 0 \), say
due to uncertainty, a quantile constraint may be envisioned

Background information: \( y = c(x, v) \)? no available formula!
there is only a database in \( (x, v, y) \)-space, \( \{ (x^k, v^k, y^k) \}_{k=1}^N \)

view the database as an empirical distribution for random
variables \( X = (X_1, \ldots, X_n), \ V = (V_1, \ldots, V_m), \ Y \)
use regression of \( Y \) on \( X_1, \ldots, X_N \) to get a function \( y = \tilde{c}(x) \)
then impose the constraint \( \tilde{c}(x) \leq 0 \) on the design \( x \)

shouldn’t the regression adapt then to the intended constraint?
Some References on Generalized Regression


Finishing the Quadrangle Scheme

\[
\begin{align*}
\text{risk } R & \leftrightarrow D \text{ deviation} \\
\text{optimization} & \quad \updownarrow S \quad \downuparrow \\
\text{regret } V & \leftrightarrow E \text{ error}
\end{align*}
\]

Error versus regret

The one-to-one correspondence

\[
\begin{align*}
\mathcal{E}(X) &= \mathcal{V}(X) - EX, \\
\mathcal{V}(X) &= EX + \mathcal{E}(X)
\end{align*}
\]

coordinates error and regret with the same “statistic”

\[
S(X) = \arg\min_C \mathcal{E}(X - C) \leftrightarrow S(X) = \arg\min_C \{C + \mathcal{V}(X - C)\}
\]

Final links: nonunique but “natural” inversions \( D \to \mathcal{E}, \ R \to \mathcal{V} \)
articulated with a scaling parameter $\lambda > 0$

$S(X) = EX$
$= \text{mean}$

$\mathcal{E}(X) = \lambda (E[X^2])^{1/2}$
$= L^2\text{-error, scaled}$

$D(X) = \lambda \sigma(X)$
$= \text{standard deviation, scaled}$

$R(X) = EX + \lambda \sigma(X)$
$= \text{safety margin risk}$

$V(X) = EX + \lambda (E[X^2])^{1/2}$
$= L^2\text{-regret}$

properties: \textit{aversity} with \textit{convexity}, but NOT coherency
The Quantile-Based Quadrangle

at any probability level \( p \in (0, 1) \)

\[
\begin{align*}
S(X) &= q_p(X) = \text{VaR}_p(X) \\
&= \text{quantile} \\
\mathcal{R}(X) &= Q_p(X) = \text{CVaR}_p(X) \\
&= \text{superquantile} \\
D(X) &= Q_p(X - EX) = \text{CVaR}_p(X - EX) \\
&= \text{superquantile deviation} \\
E(X) &= E[\frac{p}{1-p}X_+ + X_-] \\
&= \text{Koenker-Basset error, normalized} \\
\mathcal{V}(X) &= \frac{1}{1-p}E[X_+] \\
&= \text{expected absolute loss, scaled}
\end{align*}
\]

properties: **aversity with coherency**
The Median-Based Quadrangle

the quantile case at probability level \( p = 1/2 \)

\[ S(X) = \text{VaR}_{1/2}(X) = q_{1/2}(X) \]

= median

\[ R(X) = \text{CVaR}_{1/2}(X) = Q_{1/2}(X) \]

= “supermedian” (average in tail above median)

\[ D(X) = \text{CVaR}_{1/2}(X - EX) = Q_{1/2}(X - EX) \]

= supermedian deviation

\[ E(X) = E|X| \]

= \( L^1 \)-error

\[ V(X) = 2E[X_+] \]

= \( L^1 \)-regret

properties: aversity with coherency
The Max-Based Quadrangle

corresponding to the limit of the quantile case as $p \to 1$

\[ S(X) = \frac{1}{2} [\sup X + \inf X] \]
\[ = \text{center of the (essential) range of } X \]

\[ R(X) = \sup X \]
\[ = \text{top of the (essential) range of } X \]

\[ D(X) = \frac{1}{2} [\sup X - \inf X] \]
\[ = \text{radius of the (essential) range of } X \]

\[ E(X) = \sup |X| \]
\[ = \mathcal{L}^\infty \text{-error} \]

\[ V(X) = \sup [X - EX] \]
\[ = \mathcal{L}^\infty \text{-regret (max excess of “cost” over average)} \]

properties:  **aversity** with **coherency**
The Log-Exponential-Based Quadrangle

\[ S(X) = \log E[e^X] \]
\[ = \textbf{dual} \text{ expression for Boltzmann-Shannon entropy} \]
\[ R(X) = \log E[e^X] \]
\[ = \text{yes, the same as } S(X)! \]
\[ D(X) = \log E[e^X - EX] \]
\[ = \text{log-exponential deviation} \]
\[ E(X) = E[\varepsilon(X)] \text{ with } \varepsilon(x) = e^x - x - 1 > 0 \text{ when } x \neq 0 \]
\[ = \text{exponential error} \]
\[ V(X) = E[v(X)] \text{ with } v(x) = e^x - 1 \begin{cases} > 0 \text{ when } x > 0 \\ < 0 \text{ when } x < 0 \end{cases} \]
\[ = \text{exponential regret} \]

properties: \textbf{aversity with coherency}