THE FUNDAMENTAL QUADRANGLE

relating quantifications of various aspects of a random variable

---

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

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

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:

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

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

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

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

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

$=$ 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 $C \in \boldsymbol{R}$, 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

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

Quantiles and Quantile Regression

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

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

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

$\mathcal{D}(X) = \sum_{k=1}^r \lambda_k \mathrm{CVaR}_{\alpha_k}(X - EX)$ 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 $\mathcal{E}$:

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

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

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

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:

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

Risk aspects of portfolio: in objective or constraints

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

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.