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

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

Optimization context

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

Factor models

Simplication via "factors" $X_1, \dots, X_n$:

each Y_i approximated by $\hat{Y}_i = c_{i0} + c_{i1}X_1 + \cdots + c_{in}X_n$

$Y(x_1, \dots, x_m)$ replaced in optimization by $\hat{Y}(x_1, \dots, x_m)$

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, $F_X(x) = P(X \le x)$

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

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

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

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

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

$$Z_0(c_1, \ldots, c_n) = Y - [c_1 X_1 + \cdots + c_n X_n]$$

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.