$\mathcal{L}_p(\Omega, \mathcal{F}, P)$ is a complete space $\rightarrow$ Every Cauchy sequence has a limit.

Linear (vector space) + Normed + Complete $\rightarrow$ Banach space

"Theory of Linear Operators" (Banach text)

We discuss risk measures of random variables that are elements of $\mathcal{L}_p(\Omega, \mathcal{F}, P)$.

---

3 important spaces:
- finite dimensional
- Banach spaces
- space of continuous functions

$\rho : \mathcal{Z} \rightarrow \mathbb{R}$ (assigns real number to random variable of random losses)

Axioms: (R1) Convexity: $\forall Z, V \in \mathcal{Z}$, $\forall t \in [0,1]$

then $\rho(tZ + (1-t)V) \leq t\rho(Z) + (1-t)\rho(V)$

(R2) Monotonicity: if $Z, V \in \mathcal{Z}$, $Z \leq V$ ($Z(\omega) \leq V(\omega)$ with probability 1)

Then $\rho(Z) \leq \rho(V)$

(R3) Translation Equivalence: $c \in \mathbb{R}$, $Z \in \mathcal{Z}$ then

$\mathbb{1} \in \mathcal{Z}$ RV that is always equal to 1

$$\rho(Z + c\mathbb{1}) = \rho(Z) + c$$

(R4) Positive Homogenity: if $\gamma > 0$ and $Z \in \mathcal{Z}$ then

$$\rho(\gamma Z) = \gamma\rho(Z)$$

Risk measure ρ is coherent if it satisfies (R1) - (R4)

- in finite dimension, convex functions are continuous
- in infinite dimension, convex functions are not necessarily continuous
- in $\mathcal{L}_p$, functions satisfying (R1), (R2) are continuous.

proof omitted

Risk measure Examples:

① $\rho(Z) = E[Z]$ (Everything comes from Expectation being a linear operator)

convexity $E[tZ + (1-t)V] = tE[Z] + (1-t)E[V]$ (relaxation)

monotonicity $Z \preceq V \Rightarrow E[Z] \leq E[V]$

translation equivalence $E[Z + c\mathbb{1}] = E[Z] + c$

positive homogenity $E[\gamma Z] = \gamma E[Z] \quad \forall \gamma$ (relaxation)

② mean - semideviation $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty]$, $\rho: \mathcal{Z} \to \mathbb{R}$

$$\rho(Z) \doteq E[Z] + \varkappa \| (Z - E[Z])_+ \|_p, \quad \varkappa \in [0,1]$$

Suppose $\varkappa = 1$, then $\rho(Z) = E[Z] + \|(Z - E[Z])_+\|_p$

$$= E[Z] + \left(E[\underbrace{(Z - E[Z])_+^p}_{\max(0,\, Z - E[Z])}]\right)^{1/p}$$

$\Rightarrow$ Semideviation: we only consider deviations from mean that are bad.

Clearly, $\rho(Z) \geq E[Z]$

first we show $\|\cdot\|_p$ is a valid norm:

(1) $\|v+z\| \leq \|v\| + \|z\|$

(2) $\|\gamma z\| = \gamma \|z\| \quad \forall \gamma > 0$

Entropic Risk measure (popular in dynamical systems / economics)

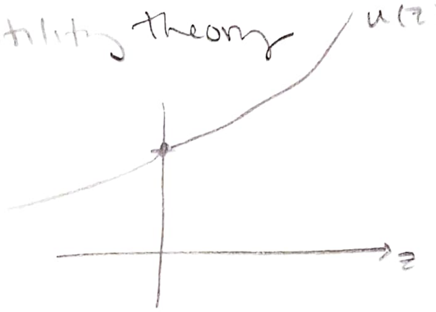
not coherent since it does not satisfy positive homogeneity

Let

$$Z \in \mathcal{Z} \Leftrightarrow E[e^{k|Z|}] < +\infty \quad \forall\, k > 0$$

$$\rho(Z) = \frac{1}{k} \ln(E[e^{kZ}])$$

Connection to Von Neumann Utility theory



$$Z \rightarrow u(Z) = e^{kZ}$$

$$E[u(Z)] = E[e^{kZ}]$$

$$\rho(Z) = u^{-1}(E[u(Z)])$$
$$= \frac{1}{k} \ln(E[e^{kZ}])$$

Then

$$e^{k\rho} = E[e^{kZ}]$$

Convexity of $\rho(Z)$

$$\rho(\alpha Z + (1-\alpha)V) = \frac{1}{k} \ln\left(E\left[\exp(\alpha k Z + (1-\alpha) k V)\right]\right)$$
$$= \frac{1}{k} \ln\left(E[(e^{kZ})^{\alpha} (e^{kV})^{(1-\alpha)}]\right)$$

By Hölder's Inequality

$$E[(e^{kZ})^{\alpha} (e^{kV})^{1-\alpha}] \le (E[e^{kZ}])^{\alpha} (E[e^{kV}])^{1-\alpha}$$

Risk Measures 11/13/2024

1) Optimized Certainty Equivalent (prepare t to pay for expected risk)

$u: \mathbb{R} \to \mathbb{R}$ non decreasing, convex w/ $1 \in \partial u(0)$

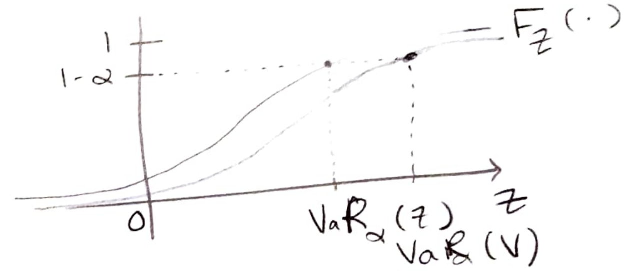
$|u(t)| \le C t^p \ \forall t \in \mathbb{R}, \ C > 0, \ p \ge 1$

$$\rho(Z) \doteq \inf_{t \in \mathbb{R}} \{ t + E[u(Z - t)] \}$$

ρ is convex measure of risk and coherent if u is positive homogeneous.

$V \le Z$

2) Value at Risk



at level $\alpha \in (0,1)$, $Z \in \mathcal{Z}$

VaR is the $(1-\alpha)$ quantile of distribution of Z

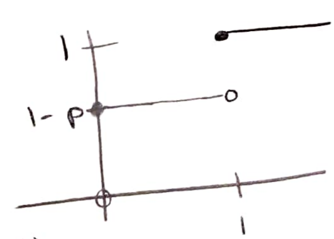
$P[Z > VaR_\alpha] \le \alpha$

$$VaR_\alpha(Z) \doteq \inf\{\eta : F_Z(\eta) \ge 1-\alpha\} = F_Z^{-1}(1-\alpha)$$

VaR is monotonic, has translation property, and positive homogeneous but it is not convex.

Example: $Z \sim Bern(p)$ $\quad P(Z=0) = 1-p$, $\ P(Z=1) = p$

$V \perp\!\!\!\perp Z$ with same dist.



For $p < \alpha < 1$ we have $VaR_\alpha(Z) = VaR_\alpha(V) = 0$

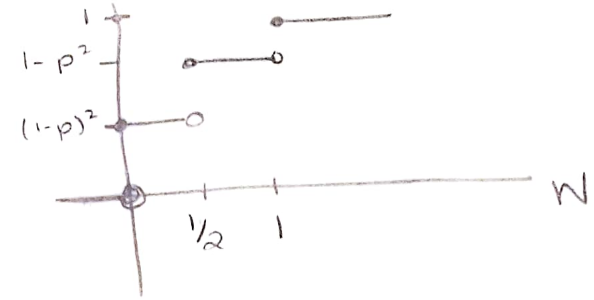
For $p < \alpha < 1-(1-p)^2$, $\quad VaR_\alpha(\lambda Z + (1-\lambda)V) > 0 = \lambda VaR_\alpha(Z) + (1-\lambda) VaR_\alpha(V)$

$$VaR_\alpha(Z) = \begin{cases} 1 & \text{if } \alpha < p \\ 0 & \text{if } \alpha \ge p \end{cases}$$

$\forall \lambda \in (0,1)$

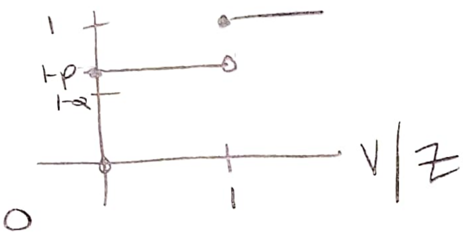
contradicting convexity

$W = \frac{1}{2} Z + \frac{1}{2} V$



We may choose α:

$p < \alpha < 1-(1-p)^2$



$VaR_\alpha(V) = VaR_\alpha(Z) = 0$

$VaR_\alpha(W) = \frac{1}{2}$ $\Rightarrow$ does not encourage diversification since risk is not reduced by combining uncorrelated risks.

VaR prefers large losses with small probability (prefers Z or V over W)

3) Average Value at Risk at level $\alpha \in (0,1]$ is

$$AVaR^+_\alpha(Z) \doteq \frac{1}{\alpha} \int_0^\alpha VaR_\beta(Z)\, d\beta$$

if the $(1-\alpha)$ quantile of Z is unique, then $P[Z \geq VaR_\alpha(Z)] = \alpha$ and we change variables to obtain

$$AVaR_\alpha(Z) = \frac{1}{\alpha} \int_{VaR_\alpha(Z)}^{1} z\, dF_Z(z) = E[Z \mid Z \geq VaR_\alpha(Z)]$$

So it is also called the Conditional Value at Risk

with assumption that $(1-\alpha)$-quartile is unique

G - profits

$$AVaR_\alpha^-(G) \doteq AVaR_\alpha^+(-G)$$

$$= \inf_\eta \left\{ \eta + \frac{1}{\alpha} E[(-G-\eta)_+] \right\}$$

$t = -\eta$

$$= \inf_t \left\{ -t + \frac{1}{\alpha} E[(t-G)_+] \right\}$$

---

minimize

Risk measure for Discrete RV (Next Homework)

Z RV : $P(Z = z_i) = p_i$ for $i = 1, \dots, K$

$p=1$: $\rho(Z) = E[Z] + \varkappa E[(Z - E[Z])_+]$ (first order semideviation)

Let $\mu = E[Z] = \frac{1}{K}\sum_{j=1}^{K} p_j z_j$ then

$$\rho(Z) = \mu + \varkappa \sum_{j=1}^{K} [p_j \underbrace{(z_j - \mu)_+}_{v_j}]$$

$v_j \leftarrow \max(0, z_j - \mu)$

To minimize $\rho(z)$ we can write:

$$\min_v \; \mu + \varkappa \sum_{j=1}^{K} p_j v_j$$

$$\text{s.t. } v_j \geq z_j - \mu \quad j = 1, \dots, K$$

$$v_j \geq 0 \quad j = 1, \dots, K$$

2 Stage Problem: $\min \; c^T x + \rho(Q(x))$

s.t. $x \in X$

$W_j = \underline{Q}_j(x)$ - Lower bound for $Q_j(x)$ (multicut method)

$Q_j(x) \geq W_j$ $\rightarrow$ calculate $\rho(W)$

For Bern Example

$$u(z) = (z-\eta)_+$$

$$AVaR_\alpha(z) = \begin{cases} \frac{1}{\alpha}\int_0^\alpha 1 \cdot d\beta = 1 & \alpha < p \\ \frac{1}{\alpha}\int_0^p 1 d\beta = \frac{p}{\alpha} & \alpha \geq p \end{cases}$$

Extremal Representation: for $\alpha \in [0,1]$ we have

$$AVaR_\alpha(z) = \inf_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{\alpha} E[(z-\eta)_+] \right\}$$

proof: we have

$$F^{(-2)}(z,1) = AVaR_1(z) = \int_0^1 F_z^{-1}(\beta)\, d\beta = E[z]$$

$$AVaR_\alpha(z) = \frac{1}{\alpha}\int_0^\alpha VaR_\gamma(z)\, d\gamma$$

$$= \frac{1}{\alpha}\int_0^\alpha F_z^{-1}(1-\gamma)\, d\gamma \qquad \beta = 1-\gamma$$

$$= \frac{1}{\alpha}\int_1^{1-\alpha} F_z^{-1}(\beta)(-d\beta)$$

$$= \frac{1}{\alpha}\int_{1-\alpha}^1 F_z^{-1}(\beta)\, d\beta = \frac{1}{\alpha}\left(E[z] - F_z^{(-2)}(1-\alpha)\right)$$

substituting $F_z^{(-2)}(\alpha) = \sup_{\eta \in \mathbb{R}} \{\alpha\eta - E[(\eta - z)_+]\}$ we obtain
(does not require unique $(1-\alpha)$-quantile) (conjugate duality of $F^{(2)}$ and $F^{(-2)}$)

$$AVaR_\alpha(z) = \frac{1}{\alpha}E[z] - \frac{1}{\alpha}\sup_{\eta \in \mathbb{R}}\left\{(1-\alpha)\eta - E[(\eta - z)_+]\right\}$$

$$= \frac{1}{\alpha}\inf_{\eta \in \mathbb{R}}\left\{-(1-\alpha)\eta + E[z] - E[(\eta - z)_+]\right\}$$

observe $z + (\eta - z)_+ = \eta + (z-\eta)_+$ we get

$$AVaR_\alpha(z) = \inf_{\eta \in \mathbb{R}}\left\{\eta + \frac{1}{\alpha}E[(z-\eta)_+]\right\}$$

$$\text{AVaR}_\alpha^+(Z) = \min_\eta \left\{ \eta + \frac{1}{\alpha} E[(Z-\eta)_+] \right\}$$

$$= \min_\eta \left\{ \eta + \frac{1}{\alpha} \sum_{k=1}^{K} p_k (z_k - \eta)_+ \right\}$$

$$= \min_{\eta, v} \left\{ \eta + \frac{1}{\alpha} \sum_{k=1}^{K} p_k v_k \right\}$$

$$\text{s.t. } v_k \geq z_k - \eta \qquad k = 1, \dots, K$$

$$v_k \geq 0$$

AVaR is bad risk measure for Garden State example b/c its too conservative. we may use

$$\min \ (1-\varphi) E[Z] + \varphi \, \text{AVaR}_\alpha^+(Z)$$

Convexity of $\text{AVaR}_\alpha^+(\cdot)$

$$\text{AVaR}_\alpha(Z) = \min_\eta \left\{ \eta + \frac{1}{\alpha} E[(Z-\eta)_+] \right\} \qquad \hat{\eta}_Z \text{ - optimal}$$

$$\text{AVaR}_\alpha(V) = \min_\eta \left\{ \eta + \frac{1}{\alpha} E[(V-\eta)_+] \right\} \qquad \hat{\eta}_V \text{ - optimal}$$

$$\text{AVaR}_\alpha(\lambda Z + (1-\lambda) V) = \min_\eta \left\{ \eta + \frac{1}{\alpha} E[(\lambda Z + (1-\lambda) V - \eta)_+] \right\}$$

choose $\eta = \lambda \hat{\eta}_Z + (1-\lambda)\hat{\eta}_V$

$$\leq \lambda \hat{\eta}_Z + (1-\lambda)\hat{\eta}_V + \frac{1}{\alpha} E\left[(\lambda Z + (1-\lambda)V - (\lambda \hat{\eta}_Z + (1-\lambda)\hat{\eta}_V))_+\right]$$

$$= \lambda \hat{\eta}_Z + (1-\lambda)\hat{\eta}_V + \frac{1}{\alpha} E\left[(\lambda(Z-\hat{\eta}_Z) + (1-\lambda)(V-\hat{\eta}_V))_+\right] \qquad \boxed{(\cdot)_+ \text{ is convex}}$$

$$\leq \lambda \left(\hat{\eta}_Z + \frac{1}{\alpha} E[(Z-\hat{\eta}_Z)_+]\right) + (1-\lambda)\left(\hat{\eta}_V + \frac{1}{\alpha} E[(V-\hat{\eta}_V)_+]\right)$$

$$= \lambda \, \text{AVaR}_\alpha^+(Z) + (1-\lambda) \, \text{AVaR}_\alpha^+(V)$$

$$\min_x \; c^T x + \rho(w) = \min_{x,v} \; c^T x + \mu + \mathcal{H} \sum_{j=1}^{K} p_j v_j$$

$$v_j \geq w_j - \mu \qquad (*)$$

$$v_j \geq 0$$

$$\mu = \sum_{j=1}^{K} p_j w_j$$

w_j - cost of scenario j → defined by cuts

v_j - upper bound that follows from representation of risk measure

$\min \; c^T x + E[Q(x)]$

$Q(x)$ has realizations $Q_j(x)$ $j=1,\dots,K$ with prob. p_j

multicut Algo:

master problem $\min \; c^T x + \boxed{\sum_k p_k w_k}$ → is lower bound for $E[Q(x)]$

$w_k \geq$ cuts for scenario $k=1,\dots,K$

$x \in X$

Now we want ~~to~~ solve $\min \; c^T x + \rho(Q(x))$

$\rho(w)$ is a lower bound of $\rho(Q(x))$

$\rho(w)$ has a LP representation that can be absorbed into the master problem

see constraints in $(*)$.

~~Typically min of 2 convex functions is not~~ convex

~~But~~ since $\eta + \frac{1}{\alpha} E[(Z-\eta)_+]$ is jointly convex in (η, Z) we have $\min_{\eta} \{\eta + \frac{1}{\alpha} E[(Z-\eta)_+]\}$

is convex.

Additional Thm: $f(x,y)$ jointly convex in (x,y)

Then $F(x) = \min_{y} \{f(x,y)\}$ is convex (in x)

* Can prove for extra credit.

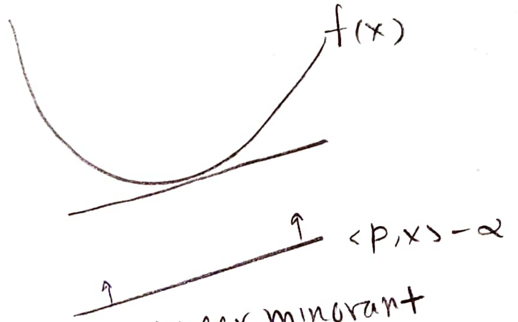
Other properties for AVaR follows directly from

$$\mathrm{AVaR}_{\alpha}(Z) = \int_{0} \mathrm{VaR}_{\beta}(Z)\, d\beta$$

so AVaR is coherent risk measure.

Conjugate Duality for Risk Measures

Recall convex conjugate functions



$f(x) \geq \langle p, x\rangle - \alpha \quad \forall x$

$\alpha \geq \langle p, x\rangle - f(x) \quad \forall x$

$\alpha^* \doteq \sup_{x} \{\langle p, x\rangle - f(x)\}$

(smallest possible α is largest RHS)

$f^*(p) \doteq \sup_{x} \{\langle p, x\rangle - f(x)\}$

$$f(x) = \sup_{p} \{\langle p, x\rangle - f^*(p)\} = f^{**}(x)$$

(Werner Fenchel Typed Lectures on this topic available at Princeton)

$\rho(z)$ measure of risk that is convex and finite

$$\rho(z) < +\infty \quad \forall\, z \in \mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$$

$1 \le p < \infty$

$$\rho^*(\zeta) = \sup_{\mathcal{Z}} \{ \underbrace{\langle \zeta, z \rangle}_{} - \rho(z) \}$$

· linear continuous functional on $\mathcal{Z}$ which must have form:

$$\langle \zeta, z \rangle = \int_\Omega z(\omega)\zeta(\omega)P(d\omega) \qquad (*)$$

In finite dimension

$$\langle c, x \rangle = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

$$\langle c, x \rangle_p = p_1c_1x_1 + p_2c_2x_2 + \dots + p_nc_nx_n = \sum_{i=1}^n p_ic_ix_i$$

where $\sum_{i=1}^n p_i = 1$, $p_j \ge 0$

$\langle \cdot, \cdot \rangle_p$ prevents very large dimension to blow up $\to \infty$ as $\langle \cdot, \cdot \rangle$ does.

Let $n \to \infty$ then $\langle c, x \rangle < \infty$ if $\sum_{i=1}^\infty c_i^2 < +\infty$, $\sum_{i=1}^\infty x_i^2 < +\infty$

then we recover the form $(*)$

Let $\mathcal{Z} = \mathcal{L}_2(\Omega, \mathcal{F}, P)$ then $\forall\, z, \zeta \in \mathcal{Z}$:

$$\int_\Omega z^2(\omega)P(d\omega) < +\infty, \quad \int_\Omega \zeta^2(\omega)P(d\omega) < +\infty$$

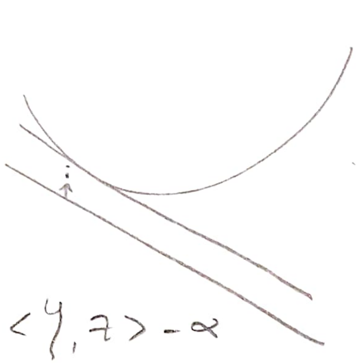
So

$$\int_\Omega z(\omega)\zeta(\omega)P(d\omega) \le \frac{1}{2}\int_\Omega (z^2(\omega) + \zeta(\omega))P(d\omega) < +\infty$$

major question in functional Analysis was the representation of all linear functionals which is

$$\langle z, \xi \rangle = \int_{\Omega} z(\omega)\xi(\omega)P(d\omega)$$

$$\alpha^* = \rho^*(\xi) = \sup_{z \in \mathcal{Z}} \{\langle z, \xi \rangle - \rho(z)\}$$



Biconjugate function of ρ

$$\rho^{**}(z) = \sup_{\xi \in \mathcal{Z}^*} \{\langle z, \xi \rangle - \rho^*(\xi)\}, \quad z \in \mathcal{Z}$$

Fenchel Moreau Thm: if ρ is proper, convex and lower semicontinuous then $\rho^{**} = \rho$

Moreau - extend Fenchel theory to infinite dimensional spaces.

$\mathcal{Z}^*$ is the space where all linear continuous functionals exist. (Dual space)

Dual Representation of Coherent measures of Risk (Thm)

if $\rho: \mathcal{Z} \to \bar{\mathbb{R}}$ is convex, proper, LSC, then $\rho(z) = \rho^{**}(z)$

(i) monotonicity condition is satisfied $\Leftrightarrow$ every $\xi \in \mathcal{A}$ is nonnegative (where $\mathcal{A} = \operatorname{dom} \rho^*$)

(ii) translation condition satisfied $\Leftrightarrow$

$$\int_{\Omega} \xi(\omega)P(d\omega) = 1, \quad \forall \xi \in \mathcal{A}$$

(iii) positive homogenity satisfied $\Leftrightarrow$ $\rho(\cdot)$ can be represented in the form $\rho(z) = \sup_{\xi \in \mathcal{A}} \{\langle z, \xi \rangle\} \ \forall z \in \mathcal{Z}$

For (ii), ($\Rightarrow$)

$$\rho^*(\xi) = \sup_z \{ \langle \xi, z \rangle - \rho(z) \}$$

$$= \sup_{z,c} \{ \langle \xi, z + c \rangle - \rho(z+c) \}$$

$$= \sup_{z,c} \{ \langle \xi, z \rangle + \langle \xi, c\mathbb{1} \rangle - \rho(z) - c \}$$

$$= \sup_z \{ \langle \xi, z \rangle - \rho(z) \} + \sup_c \{ \langle \xi, c\mathbb{1} \rangle - c \}$$

$$= \rho^*(\xi) + \sup_c \{ \langle \xi, c\mathbb{1} \rangle - c \}$$

So $\sup_c \{ \langle \xi, c\mathbb{1} \rangle - c \} = 0$ only possible if

$$\langle \xi, \mathbb{1} \rangle = 1. \quad \square$$

By (i), (ii) ξ is a probability density!

Risk measures 11/20/24

Recall convex conjugate functions

- $\rho(z)$ convex
- $\rho^*(\zeta) = \sup_{z \in \mathcal{Z}} \{ \langle \zeta, z \rangle - \rho(z) \}$
- $\rho(z) = \sup_{\zeta \in A} \{ \langle \zeta, z \rangle - \rho^*(\zeta) \}$

where $A = \operatorname{dom} \rho^*$

For (i), ($\Rightarrow$) Suppose ρ satisfies montonicity condition, but $\exists\ \zeta \in A : \mathbb{P}[\zeta < 0] > 0$. Then $\langle \mathbb{1}_{\{\zeta < 0\}}, \zeta \rangle < 0$

For any Z_0 such that $\rho(Z_0) < +\infty$ consider the ray $Z_t = Z_0 - t\,\mathbb{1}_{\{\zeta < 0\}}$. By construction, $Z_0 \geq Z_t \ \forall\ t \geq 0$ and by monotonicity condition $\rho(Z_0) \geq \rho(Z_t)\ \forall t \geq 0$.

$$\rho^*(\zeta) \geq \sup_{t \geq 0} \{ \langle Z_t, \zeta \rangle - \rho(Z_t) \}$$

$$\geq \sup_{t \geq 0} \{ \langle Z_0, \zeta \rangle - t \langle \mathbb{1}_{\{\zeta < 0\}}, \zeta \rangle - \rho(Z_0) \}$$

$$= +\infty$$

which contradicts assumption $\zeta \in A$. Therefore $\mathbb{P}[\zeta < 0] = 0$ $\forall\ \zeta \in A$.

($\Leftarrow$) Suppose every element of A is non negative. Let $Z \leq V$, then $\langle Z, \zeta \rangle \leq \langle V, \zeta \rangle \ \forall\ \zeta \in A$. By (dual) definition of $\rho(\cdot)$, in

$$\rho(Z) \leq \rho(V)$$

(iii) ($\Rightarrow$) Let ρ be positively homogeneous. Let $\gamma > 0$.

$$\rho^*(\zeta) = \sup_{z \in Z} \{ \langle z, \zeta \rangle - \rho(z) \}$$

$$= \sup_{z \in Z} \{ \langle \gamma z, \zeta \rangle - \rho(\gamma z) \}$$

$$= \sup_{z \in Z} \{ \gamma \langle z, \zeta \rangle - \gamma \rho(z) \}$$

$$= \gamma \sup_{z \in Z} \{ \langle z, \zeta \rangle - \rho(z) \}$$

$$= \gamma \rho^*(\zeta) = 0$$

Plugging into dual representation

$$\rho(z) = \sup_{\zeta \in \mathcal{A}} \{ \langle z, \zeta \rangle - \rho^*(\zeta) \}$$

$$= \sup_{\zeta \in \mathcal{A}} \{ \langle z, \zeta \rangle \}$$

($\Leftarrow$) if $\rho(z) = \sup_{\zeta \in \mathcal{A}} \{ \langle z, \zeta \rangle \}$ then its positively homogen.

$$\sum_{k=1}^{K} \mu_k = \sum_{k=1}^{K} \lambda_k + \left(\sum_{k=1}^{K} p_k\right)\lambda_0 = 1 \quad \text{by constraints}$$

$$\mu_k = \lambda_k + p_k\Big(1 - \underbrace{\sum_{k=1}^{K} \lambda_k}_{\sum_{k=1}^{K} \alpha p_k}\Big) \geq \lambda_k + p_k\Big(1 - \alpha \underbrace{\sum_{k=1}^{K} p_k}_{1}\Big)$$

$$= \lambda_k + p_k(1-\alpha) \geq 0$$

So μ_k are probabilities. Then

$$\rho(Z) = \max_{\frac{d\mu}{dP} \in \mathcal{A}} E^{\mu}[Z] \qquad \text{(Dual representation)}$$

Let $\rho: \mathcal{Z} \to \overline{\mathbb{R}}$ be LSC, proper, coherent risk measure, then

(i) $\partial\rho(Z)$ is a nonempty, bounded, weakly compact subset of $\mathcal{Z}^*$ given by

$$\partial\rho(Z) = \arg\max_{\zeta \in \partial\rho(0)} \{\langle \zeta, Z \rangle\}$$

(ii) ρ is Hadamard directionally differentiable at Z and

$$\rho'(H, Z) = \sup_{\zeta \in \partial\rho(Z)} \{\langle \zeta, H \rangle\} \quad \forall H \in \mathcal{Z}.$$

Dual Representation Examples

$p=1$. $\Omega = \{\omega_1, \dots, \omega_K\}$ with probabilities $p_1, \dots, p_K$.

$z_k = Z(\omega_k)$ for $k = 1, \dots, K$

mean semideviation

$$\rho(Z) = E[Z] + \varkappa E[(Z - E[Z])_+]$$

$$\mu = \sum_{k=1}^{K} p_k z_k$$

$$= \mu + \varkappa \sum_{k=1}^{K} p_k \underbrace{(z_k - \mu)_+}_{v_k}$$

$$v_k \doteq z_k - \mu$$

$$v_k \geq 0.$$

$$\rho(Z) = \min \quad \mu + \varkappa \sum_{k=1}^{K} p_k v_k$$

$$\text{s.t.} \quad \mu = \sum_{k=1}^{K} p_k z_k \qquad \leftarrow (\lambda_0)$$

$$\mu + v_k \geq z_k \quad k = 1, \dots, K \qquad \leftarrow (\lambda_k \geq 0)$$

$$v_k \geq 0 \qquad k = 1, \dots, K$$

Dual problem

$$\max \quad \lambda_0 \sum_{k=1}^{K} p_k z_k + \sum_{k=1}^{K} \lambda_k z_k$$

$$\text{s.t.} \quad \lambda_0 + \sum_{k=1}^{K} \lambda_k = 1$$

$$\lambda_k \leq \varkappa p_k \quad k = 1, \dots, K$$

$$\lambda_k \geq 0 \quad k = 1, \dots, K$$

$$\max \sum_{k=1}^{K} \mu_k z_k \quad \text{s.t. same constraints} \qquad \text{where} \quad \mu_k = \lambda_k + p_k \lambda_0$$

For a mapping $F: \mathbb{R}^n \to \mathcal{Z}$, denote $[F(x)](\omega) = f_\omega(x)$

and consider the composite function $\phi(\cdot) \doteq \rho(F(\cdot))$

Lemma Let $F: \mathbb{R}^n \to \mathcal{Z}$ be a convex mapping, that is,

$$F(\lambda x + (1-\lambda)y) \preceq \lambda F(x) + (1-\lambda)F(y) \text{ a.s. } \forall \lambda \in [0,1],$$

then F is continuous, directionally differentiable with directional derivative at $\bar{x} \in \mathbb{R}^n$

$$[F'(\bar{x}, h)](\omega) = f'_\omega(\bar{x}, h), \quad \omega \in \Omega, \ h \in \mathbb{R}^n$$

"derivative of random function is derivative

Furthermore, $\phi(\cdot) = \rho(F(\cdot))$ is convex.

Extension of Strassen's Thm

Suppose $F: \mathbb{R}^n \to \mathcal{Z}$ is convex and ρ is finite valued, coherent

Then $\phi = \rho \circ F$ is directionally differentiable and subdifferentiable at any $\bar{x}$, it has finite valued directional derivative $\forall h \in \mathbb{R}^n$ and (Let $\bar{z} = F(\bar{x})$)

$$\rho'(\bar{z}, F'(\bar{x}, h)) = \sup_{\zeta \in \partial\rho(\bar{z})} \left\{ \int_\Omega f'_\omega(\bar{x}, h) \zeta(\omega) P(d\omega) \right.$$

$$\partial\phi(\bar{x}) = \bigcup_{\zeta \in \partial\rho(\bar{z})} \int_\Omega \partial f_\omega(\bar{x}) \zeta(\omega) P(d\omega)$$

Law Invariant Risk Measures

Defn: $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is law-invariant w.r.t. the reference probability measure P, if $\forall\, Z, V \in \mathcal{Z}$ if

$$\{Z \overset{D}{\sim} V\} \Rightarrow \{\rho(Z) = \rho(V)\}$$

This class contains all interesting risk measures (any measure calculated w/ moments)

Consider $\Omega = \{\omega_1, \dots, \omega_n\}$ with $P(\omega_i) = 1/n$ for $i = 1, \dots, n$. For a RV on Ω we can obtain the quantile function F_Z^{-1} by finding permutation $T_Z : \Omega \to \Omega$:

$$Z(T_Z(\omega_1)) \le Z(T_Z(\omega_2)) \le \dots \le Z(T_Z(\omega_n))$$

and we set $F_Z^{-1}(p) = Z(T_Z(\omega_i))$ whenever $\frac{i-1}{n} < p \le \frac{i}{n}$ for $i = 1, \dots, n$. Since 2 RVs, Z and V, have same distribution if their quantile functions coincide, 2 permutations can be found, T_Z and T_V, such that

$$Z(T_Z(\omega_i)) = V(T_V(\omega_i))$$

Thus for $T = T_V T_Z^{-1}$ we get

$$Z(\omega_i) = V(T(\omega_i)) \quad i = 1, \dots, n$$

Kusuoka Representation

Let $\Omega = \{w_1, \ldots, w_n\}$ with $P[w_i] = 1/n$. A convex risk measure $\rho: \mathcal{Z} \to \mathbb{R}$ is law-invariant if and only if a set Λ of probability vectors in $\mathbb{R}^n$ exists such that

$$\rho(z) = \sup_{x \in \Lambda} \left\{ \sum_{i=1}^{n} x_i \, AV@R_{i/n}(z) \right\} \quad \forall z \in \mathcal{Z}$$

proof: By dual representation

where $z_i = z(w_i)$, $\xi_i = \xi(w_i)$

$$\rho(z) = \max_{\xi \in \mathcal{A}} \left\{ \frac{1}{n} \sum_{i=1}^{n} z_i \xi_i \right\}, \quad z \in \mathcal{Z}$$

Denote $z_{[1]} \leq z_{[2]} \leq \cdots \leq z_{[n]}$ ordered realizations of z

By Law Invariance of ρ we have

$$\rho(z) = \max_{\xi \in \mathcal{A}} \left\{ \frac{1}{n} \sum_{i=1}^{n} z_{[i]} \xi_i \right\}, \quad z \in \mathcal{Z}$$

Let $\xi \in \mathcal{A}$. $\rho^*(\xi) = \max_{z \in \mathcal{Z}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \xi_i z_i - \rho(z) \right\}$

$$= \max_{z \in \mathcal{Z}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \xi_i z_{T(i)} - \rho(z) \right\}$$

by Law Invariance ($\rho(z \circ T) = \rho(z)$)

$$= \max_{z \in \mathcal{Z}} \left\{ \frac{1}{n} \sum_{j=1}^{n} \xi_{T^{-1}(j)} z_j - \rho(z) \right\}$$

$$= \rho^*(\xi \circ T^{-1})$$

So $\mathcal{A}$ is closed under permutation. Then

By Littlewood-Hardy-Polya inequality we have

$$\sum_{i=1}^{n} z_{[i]} \xi_i \leq \sum_{i=1}^{n} z_{[i]} \xi_{[i]}$$

Then $\rho(Z) = \max_{\zeta \in \mathfrak{A}} \left\{ \frac{1}{n} \sum_{i=1}^{n} z_{[i]} \zeta_{[i]} \right\}, \quad Z \in \mathcal{Z}$

Denote $\zeta_{[0]} = 0$. We may rewrite the sum (integration by parts)

$$\frac{1}{n} \sum_{i=1}^{n} z_{[i]} \zeta_{[i]} = \frac{1}{n} \sum_{i=1}^{n} (\zeta_{[i]} - \zeta_{[i-1]}) \sum_{j=i}^{n} z_{[j]}$$

$$= \sum_{i=1}^{n} \frac{n-i+1}{n} (\zeta_{[i]} - \zeta_{[i-1]}) \frac{1}{n-i+1} \sum_{j=i}^{n} z_{[j]}$$

Define $\lambda_i = \frac{n-i+1}{n} (\zeta_{[i]} - \zeta_{[i-1]}), \quad i = 1, \dots, n$

$$\rho(Z) = \max_{\zeta \in \mathfrak{A}} \Big\{ \sum_{i=1}^{n} \lambda_i \underbrace{\frac{1}{n-i+1} \sum_{j=i}^{n} z_{[j]}}_{\mathrm{AVaR}_{\alpha_i}(Z) \text{ with } \alpha_i = \frac{1}{n-i+1}} \Big\}$$

We see $\lambda_i \geq 0$ and

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} \frac{n-i+1}{n} (\zeta_{[i]} - \zeta_{[i-1]}) = \sum_{i=1}^{n} \frac{1}{n} \zeta_{[i]} = 1$$

■

Thm: Let $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$ where Ω is standard and atomless. Then a proper LSC convex risk measure $\rho : \mathcal{Z} \to \overline{\mathbb{R}}$ is law invariant if and only if a set Λ of probability measures on the interval $(0, 1]$ exists such that

$$\rho(Z) = \sup_{\lambda \in \Lambda} \left\{ \int_0^1 \mathrm{AVaR}_\alpha(Z) \, \lambda(d\alpha) \right\} \quad \forall Z \in \mathcal{Z}$$

Consistency with increasing convex order

$$\eta + \frac{1}{\alpha} E[(Z_1 - \eta)_+] \le \eta + \frac{1}{\alpha} E[(Z_2 - \eta)_+] \quad \forall \eta, \ \alpha \in (0,1]$$

$$\Rightarrow \min_{\eta} \left\{ \eta + \frac{1}{\alpha} E[(Z_1 - \eta)_+] \right\} \le \min_{\eta} \left\{ \eta + \frac{1}{\alpha} E[(Z_2 - \eta)_+] \right\}$$

$$\Rightarrow \quad AVaR_\alpha(Z_1) \le AVaR_\alpha(Z_2) \quad \forall \alpha$$

$$\Rightarrow \int_0^1 AVaR_\alpha(Z_1)\lambda(d\alpha) \le \int_0^1 AVaR_\alpha(Z_2)\lambda(d\alpha)$$

$$\Rightarrow \sup_{\lambda \in \Lambda} \left\{ \int_0^1 AVaR_\alpha(Z_1)\lambda(d\alpha) \right\} \le \sup_{\lambda \in \Lambda} \left\{ \int_0^1 AVaR_\alpha(Z_2)\lambda(d\alpha) \right\}$$

$$\Rightarrow \quad \rho(Z_1) \le \rho(Z_2)$$

12/4/24 Conditional Risk Mappings / Dynamical Models

Sample space Ω with sigma algebras $\mathcal{F}_1 \subset \mathcal{F}_2$ and a probability measure P on $(\Omega, \mathcal{F}_2)$. Consider spaces

$$\mathcal{Z}_1 = \mathcal{L}_p(\Omega, \mathcal{F}_1, P), \quad \mathcal{Z}_2 = \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$$

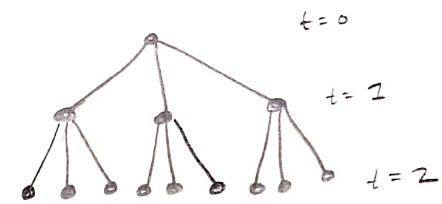
observation $t=1$ $\quad \underbrace{\mathcal{F}_1}_{} \subset \mathcal{F}_2$

events that can be distinguished at stage $t=1$ (context)

- $\rho(Z)$ may be $\mathcal{F}_1$ measurable

$\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$

$\mathcal{Z}_1$



$Z \in \mathcal{Z}_2$

Z-random cost paid at second stage

ρ is a coherent conditional risk mapping if it satisfies the following

R1) Convexity: $\rho(\alpha Z + (1-\alpha)V) \preceq \alpha \rho(Z) + (1-\alpha)\rho(V)$

$$\forall Z, V \in \mathcal{Z}_2, \ \forall \alpha \in [0,1]$$

R2) monotonicity: if $V, Z \in \mathcal{Z}_2$ and $Z \preceq V$, then $\rho(Z) \preceq \rho(V)$

R3) Translation Equivariance: if $Y \in \mathcal{Z}_1$ and $Z \in \mathcal{Z}_2$, then $\rho(Z+Y) = \rho(Z) + Y$

R4) Positive Homogeneity: if $\gamma \geq 0$ and $Z \in \mathcal{Z}_2$ then

$$\rho(\gamma Z) = \gamma \rho(Z)$$

where "$\preceq$" are understood in the P-almost sure sense. (pointwise on ω)

A mapping $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$ satisfying (R1) - (R3) is a convex conditional risk mapping.

1) Conditional Expectation

The mapping $E[\cdot \mid \mathcal{F}_1] : \mathcal{Z}_2 \to \mathcal{Z}_1$ is a conditional risk mapping

- For any $p \geq 1$ and $Z \in \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$ Jensen's inequality implies $E[|Z|^p \mid \mathcal{F}_1] \geq |E[Z \mid \mathcal{F}_1]|^p$ hence

$$\int_\Omega |E[Z \mid \mathcal{F}_1]|^p dP \leq \int_\Omega E[|Z|^p \mid \mathcal{F}_1] dP = E[|Z|^p] < +\infty$$

Thus $E[\cdot \mid \mathcal{F}_1]$ maps $\mathcal{Z}_2$ to $\mathcal{Z}_1$. Conditional Expectation is a linear operator so (R1), (R4) follow. (R2), (R3) also follow for conditional Expectation.

2) Conditional AVaR

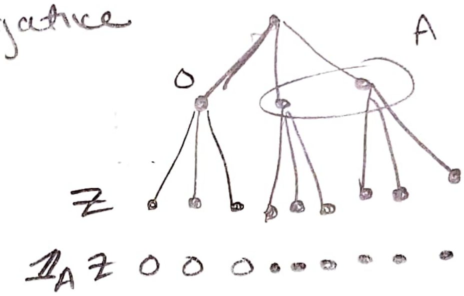
For $\alpha \in (0,1]$ define $AVaR_\alpha(\cdot \mid \mathcal{F}_1) : \mathcal{Z}_2 \to \mathcal{Z}_1$ as

$$[AVaR_\alpha(Z \mid \mathcal{F}_1)](\omega) = \inf_{Y \in \mathcal{Z}_1} \left\{ Y(\omega) + \frac{1}{\alpha} E[[Z - Y]_+ \mid \mathcal{F}_1](\omega) \right\}, \quad \omega \in \Omega$$

Local Property of Conditional Risk Mapping.

Suppose $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$ satisfies (R1) - (R4) then

$$\rho(\mathbb{1}_A Z) = \mathbb{1}_A \rho(Z) \quad \forall\ A \in \mathcal{F}_1,\ Z \in \mathcal{Z}_2$$



Proof: Consider $\rho : \mathcal{Z}_2 \to \mathcal{Z}_1$ and nonnegative function $\xi \in \mathcal{Z}_1^*$ such that

$\int_\Omega \xi(\omega)\, dP(\omega) = 1$, we associate

$$\rho_\xi(Z) \doteq \int_\Omega \xi(\omega)\, \rho(Z(\omega))\, dP(\omega), \quad Z \in \mathcal{Z}_2$$

Conditions (R1) - (R4) for ρ imply the corresponding conditions (R1) - (R4) for ρ_ξ, hence ρ_ξ is a coherent risk measure defined on $\mathcal{Z}_2 = \mathcal{L}_p(\Omega, \mathcal{F}_2, P)$. For $B \in \mathcal{F}_1$ we get by (R3)

$$\rho_\xi(Z + \alpha \mathbb{1}_B) = \int_\Omega \xi(\omega) \left[\rho(Z) + \alpha \mathbb{1}_B\right](\omega)\, dP(\omega)$$

$$= \rho_\xi(Z) + \alpha \int_B \xi(\omega)\, dP(\omega) \quad \forall\ \alpha \in \mathbb{R}$$

Since ρ_ξ is coherent it has a dual representation (see latest book for proof, using conditional expectations)

Dynamic Measures of Risk

$(Z_1, Z_2, \ldots, Z_T)$ sequence of random costs where Z_t is observed at time t. Probability space $(\Omega, \mathcal{F}, P)$ with filtration $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_T \subset \mathcal{F}$ and assume $\mathcal{F}_1 = \{\emptyset, \Omega\}$

stage wise cost

$\mathcal{Z}_t = \mathcal{L}_p(\Omega, \mathcal{F}_t, P)$, $p \in [1, \infty)$, $\mathcal{Z}_{t,T} = \mathcal{Z}_t \times \mathcal{Z}_{t+1} \times \cdots \times \mathcal{Z}_T$.

Defn A dynamic risk measure is a sequence of mappings $\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{Z}_t$, $t = 1, \ldots, T$ satisfying for all t and all $(Z_t, \ldots, Z_T)$ and $(W_t, \ldots, W_T)$ in $\mathcal{Z}_{t,T}$ the monotonicity property:

if $(Z_t, \ldots, Z_T) \preceq (W_t, \ldots, W_T)$ then

$$\rho_{t,T}(Z_t, \ldots, Z_T) \preceq \rho_{t,T}(W_t, \ldots, W_T)$$

- the value of the "tail risk measure" $\rho_{t,T}(Z_t, \ldots, Z_T)$ can be interpreted as a fair one time $\mathcal{F}_t$-measurable charge we would pay at time t instead of sequence of random costs: $Z_t, \ldots, Z_T$

Defn: A dynamic risk measure $\{\rho_{t,T}\}_{t=1}^{T}$ is called time consistent if $\forall$ $1 \le \tau < \theta \le T$ all sequences $Z, W \in \mathcal{Z}_{\tau,T}$ the conditions $Z_k = W_k$, $k = \tau, \ldots, \theta - 1$ and $\rho_{\theta,T}(Z_\theta, \ldots, Z_T) \preceq \rho_{\theta,T}(W_\theta, \ldots, W_T)$ imply that

$$\rho_{\tau,T}(Z_\tau, \ldots, Z_T) \preceq \rho_{\tau,T}(W_\tau, \ldots, W_T)$$

- Needs to be assumed, does not follow from the definition of dynamic risk measures.

For a dynamic risk measure $\{\rho_{t,T}\}_{t=1}^{T}$ we can define broader family of conditional risk measures by setting

$$\rho_{\tau,\theta}(Z_\tau,\dots,Z_\theta) = \rho_{\tau,T}(Z_\tau,\dots,Z_\theta,0,\dots,0) \quad 1 \le \tau < \theta \le T$$

Thm Suppose a dynamic risk measure $\{\rho_{t,T}\}_{t=1}^{T}$ satisfies for all $t=1,\dots,T$ and $Z_t \in \mathcal{Z}_t$ the condition:

$$\rho_{t,T}(Z_t,0,\dots,0) = Z_t$$

Then it is time consistent if and only if for all $1 \le \tau < \theta \le T$ and $Z \in \mathcal{Z}_{1,T}$ we have:

$$\rho_{\tau,T}(Z_\tau,\dots,Z_\theta,\dots,Z_T) = \rho_{\tau,\theta}(Z_\tau,\dots,Z_{\theta-1},\rho_{\theta,T}(Z_\theta,\dots,Z_T))$$

proof: Consider $Z = (Z_\tau,\dots,Z_{\theta-1},Z_\theta,Z_{\theta+1},\dots,Z_T)$

$$W = (Z_\tau,\dots,Z_{\theta-1},\rho_{\theta,T}(Z_\theta,\dots,Z_T),0,\dots,0)$$

$$\rho_{t,T}(Z_t, Z_{t+1}, \ldots, Z_T) = E[Z_t + \cdots + Z_T \mid \mathcal{F}_t] \doteq E_t[Z_t + \cdots + Z_T]$$

Tower property of expectation: $E_t[Z_t + Z_{t+1} + \cdots + Z_{T-1} + Z_T]$

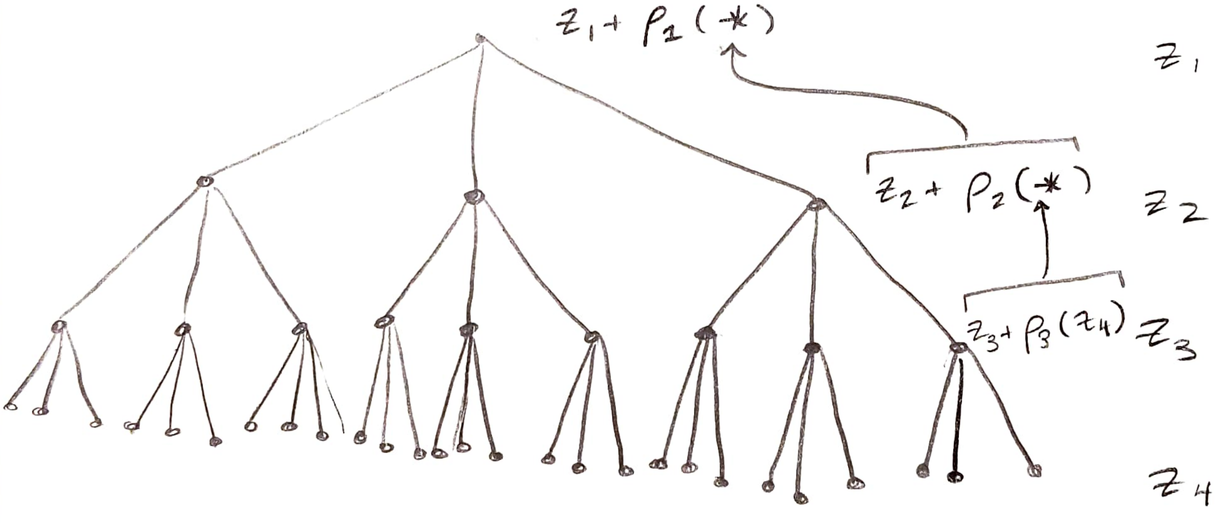
$$= E_t E_{t+1} \cdots E_{T-1}[Z_t + \cdots + Z_{T-1} + Z_T]$$

$$= E_t E_{t+1} \cdots E_{T-2}[Z_t + \cdots + Z_{T-1} + E_{T-1}[Z_T]]$$

Continuing in this way:

$=$

time consistent dynamic risk measure also has this property and is completely defined by one step dynamic risk measures ρ_t, $t = 1, \ldots, T-1$



"Composition of partially aggregated risk"

$D_1 \rightsquigarrow x_1 \rightsquigarrow D_2 \rightsquigarrow x_2 \rightsquigarrow \cdots \rightsquigarrow D_t \rightsquigarrow x_t \rightsquigarrow \cdots \rightsquigarrow D_T \rightsquigarrow x_T$

Cost $f_1(x_1) + f_2(x_2) + \dots + f_T(x_T)$

$x_t \in \mathcal{X}_t(x_{t-1}, D_{[t]})$

$Z_1 = f_1(x_1),\ Z_2 = f_2(x_2), \dots, Z_T = f_T(x_T)$ are random variables

since decision depend on random data process D_t.

Stochastic Subgradient Methods 12/11/2024

$\min_{x \in X} f(x)$ where $f: \mathbb{R}^n \to \mathbb{R}$ convex, $X \subset \mathbb{R}^n$ convex and closed. Subgradients of $f(\,)$ are not available and we use their statistical estimates

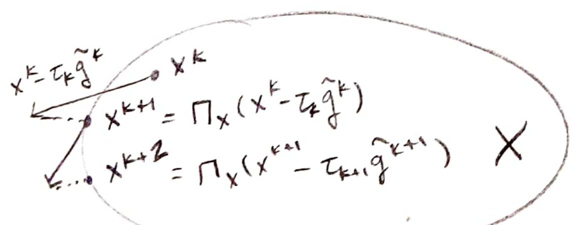
Example (EV-optimization)

$f(x) = E[F(x,Z)]$ where Z is an m-dimensional random vector, $F: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. if $F(\cdot, z)$ is convex $\forall z \in \mathbb{R}^m$ and the expected value $E[F(x,Z)]$ is finite forall x, then by Strassen's Thm $\partial f(x) = E[\partial F(x,Z)]$

Here $g \in \partial f(x)$ if and only if a measurable selector $G(x,z) \in \partial F(x,z)$ exists such that $g = E[G(x,Z)]$. At a point x, we can sample a realization $\hat{z}$ of Z and obtain a stochastic subgradient $\tilde{g}(x) = G(x,\hat{z}) \in \partial F(x,\hat{z})$. The conditional expectation of $\tilde{g}(x)$ given x is a subgradient of $f(\cdot)$ at x.

<u>method:</u> $x^{k+1} = \Pi_X(x^k - \tau_k \tilde{g}^k)$, $k = 1, 2, \dots$

where Π_X is the orthogonal projection on the set X



directions $\tilde{g}^k \in \mathbb{R}^n$ and stepsizes $\tau_k > 0$ are random thus iterates $\{x^k\}$ are random.

Assumptions $\mathcal{F}_k$ - σ-algebra generated by $\{x^1, \hat{g}^1, x^2, \dots, \hat{g}^{k-1}, x^k\}$

(A1) stepsizes τ_k are $\mathcal{F}_k$-measurable such that

(i) $\tau_k > 0$ $\quad k = 1, 2, \dots$

(ii) $\sum_{i=1}^{\infty} \tau_k = +\infty$ a.s.

(iii) $E\left[\sum_{i=1}^{\infty} \tau_k^2\right] < +\infty$

trend will dominate the error (variance)

(A2) The random vectors $\hat{g}^k$ satisfy for $k = 1, 2, \dots$ the conditions:

(i) $\hat{g}^k = g^k + e^k$ with $g^k \in \partial f(x^k)$

(ii) $E[e^k \mid \mathcal{F}_k] = 0$

(iii) $E[\|e^k\|^2 \mid \mathcal{F}_k] \le \sigma^2$, where σ^2 is a constant

(A3) The set X is compact.

Under (A3), a constant C exists, such that $\|g\| \le C$ for all $g \in \partial f(x)$ where $x \in X$.

Recall Martingale Convergence Theorem —

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and the sequence $\{\mathcal{F}_n\}$ of σ-fields satisfies $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}_n \subset \dots \subset \mathcal{F}$. A sequence of real RVs $\{X_n\}$ in $\mathcal{L}_1(\Omega, \mathcal{F}, P)$ is a martingale if

$$E[X_{n+1} \mid \mathcal{F}_n] = X_n, \quad n = 1, 2, \dots$$

super martingale $E[X_{n+1} \mid \mathcal{F}_n] \le X_n$

sub martingale $E[X_{n+1} \mid \mathcal{F}_n] \ge X_n$

Thm (Doob): if $\{X_n\}$ is a submartingale with $\sup_n E[X_n^+] < +\infty$, or a supermartingale with $\sup_n E[X_n^-] < +\infty$, then the sequence $\{X_n\}$ is convergent with probability 1 to some random limit X^∞

Example: Z scalar R.V.

$$\min_x \frac{1}{2} E[(x-z)^2] \qquad F(x,z) = \frac{1}{2}(x-z)^2$$

$$\partial_x F = \{x - z\} \Rightarrow \tilde{g}(x) = x - z$$

$$x^{k+1} = x^k - \tau_k \tilde{g}^k = x^k - \tau_k (x^k - z^k) = (1-\tau_k) x^k + \tau_k z^k$$

Choose $\tau_k = \frac{1}{k}$

$$x^{k+1} = \left(1 - \frac{1}{k}\right) x^k + \frac{1}{k} z^k \qquad x^1 = 0$$

$$= \frac{k-1}{k} x^k + \frac{1}{k} z^k$$

we see $x^k = \frac{1}{k-1} \sum_{i=1}^{k-1} z^i$ (online average)

clearly $x^k \to E[z]$ which is the optimal solution

Convergence Thm: Assume (A1) - (A3). Then the sequence $\{x^k\}$ generated by the method converges to an optimal solution with probability 1.

proof: since f is convex, we consider optimality metric of distance to optimal solution, x^*

$$\|x^{k+1} - x^*\|^2 = \|\Pi_X(x^k - \tau_k \tilde{g}^k) - x^*\|^2$$

Follow in notes. (Yuri Ermoliev)

Finite Iteration performance

Suppose we use deterministic step sizes $\tau_k > 0$, $k = 1, 2, \ldots$, we have

$$E[\|x^{k+1} - x^*\|^2 \mid \mathcal{F}_k] \le \|x^k - x^*\|^2 - 2\tau_k (f(x^k) - f(x^*)) + \hat{\sigma}^2 \tau_k^2$$

Taking full expectation

$$\tau_k E[f(x^k) - f(x^*)] \le \frac{1}{2}\left[E[\|x^k - x^*\|^2] - E[\|x^{k+1} - x^*\|^2] + \hat{\sigma}^2 \tau_k^2\right]$$

The sum of these inequalities for $k=0,1,\dots,N$ yields

$$E\left[\sum_{k=1}^{N}\tau_k\left[f(x^k)-f(x^*)\right]\right]\le\frac{1}{2}\|x^1-x^*\|^2+\frac{\hat{\sigma}^2}{2}\sum_{k=1}^{N}\tau_k$$

$$\overline{\sum_{j=1}^{N}\tau_j} \qquad \overline{\sum_{j=1}^{N}\tau_j}$$

Let $\alpha_k=\dfrac{\tau_k}{\sum_{j=1}^{N}\tau_j}>0$ we see $\sum_{k=1}^{N}\alpha_k=1$ so we may interpret them as probabilities. Then

$$E\left[\sum_{k=1}^{N}\alpha_k\left(f(x^k)-f(x^*)\right)\right]\le \text{RHS}$$

$$E\left[\sum_{k=1}^{N}\alpha_k f(x^k)\right]-f(x^*)\le \text{RHS}$$

Since f is convex we have

$$f(\bar{x}^N)\le\sum_{k=1}^{N}\alpha_k f(x^k),\quad\text{where } \bar{x}^N=\sum_{k=1}^{N}\alpha_k x^k$$

so

$$E\left[f(\bar{x}^N)\right]-f(x^*)\le\frac{1}{2}\,\frac{\|x^1-x^*\|^2+\hat{\sigma}^2\sum_{k=1}^{N}\tau_k^2}{\sum_{k=1}^{N}\tau_k}$$

We choose $\tau_k=\dfrac{\alpha}{\sqrt{N}}$ and get

$$E\left[f(\bar{x}^N)\right]-f(x^*)\le\frac{1}{2}\,\frac{\|x^1-x^*\|^2+\hat{\sigma}^2\alpha^2}{\alpha\sqrt{N}}$$

So we may choose N and α obtain a sufficient quality guarantee

For final homework: implement this method on absolute value regression. We may also use $\tau_k=\dfrac{\alpha}{\sqrt{k}}$ so we do not need to choose number of steps N. Do analysis here with this τ_k and get quality estimate.

Hint: get upper bound of $\sum_{k=1}^{N}\tau_k^2$ $\Rightarrow$ summing rectangles $\Rightarrow$ under estimate of some integral

lower bound of $\sum_{k=1}^{N}\tau_k$

Smooth Unconstrained Problems

$f(\cdot)$ is continuously differentiable and its gradient $\nabla f(\cdot)$ is Lipschitz continuous with constant L. $X = \mathbb{R}^n$

$$x^{k+1} = x^k - \tau_k \hat{g}^k, \quad k = 1, 2, \ldots$$

where $\hat{g}^k = \nabla f(x^k) + e^k$

Define $X^* = \{x \in \mathbb{R}^n : \nabla f(x) = 0\}$

Thm: Assume (A1)-(A3) and let $f^* = \min f(x)$ be finite. Then with probability 1, the sequence $\{f(x^k)\}$ is convergent and every accumulation point $\hat{x}$ of the sequence $\{x^k\}$ satisfies $\nabla f(\hat{x}) = 0$.

Averaging of Directions

Here we generate 2 sequences: the main iterated $\{x^k\}$ and the path averaged stochastic subgradients $\{z^k\}$. Given x^k, z^k we solve

$$\min_{y \in X} \left\{ \langle z^k, y - x^k \rangle + \frac{1}{2}\|y - x^k\|^2 \right\} \qquad (19)$$

denote y^k as the solution. we make a step in primal iterates

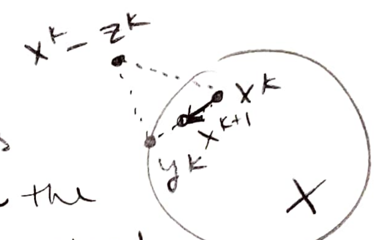
$$x^{k+1} = x^k + \tau_k (y^k - x^k)$$

we observe a stochastic subgradient $\hat{g}^{k+1}$ at x^{k+1} and update the average subgradients by

$$z^{k+1} = (1 - \alpha \tau_k) z^k + \alpha \tau_k \hat{g}^{k+1}$$

The subproblem (19) is equivalent to $y^k = \Pi_X (x^k - z^k)$

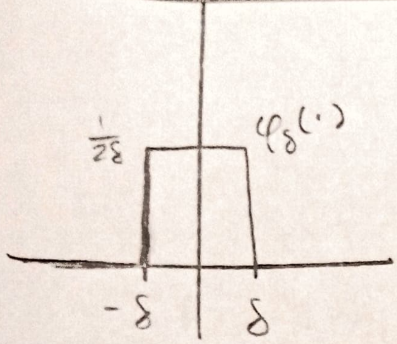
- In unconstrained case $(X = \mathbb{R}^n)$ we have $y^k = x^k - z^k$ so $x^{k+1} = x^k - \tau_k z^k$ which is analogous to stochastic subgradient method, but with the "filtered" stochastic subgradient z^k used instead of last observed $\hat{g}^k$.



Home work: $\min_x E\Big[\underbrace{|x_0 + x_1 A_1 + x_2 A_2 + A_3|}_{F(x,A)}\Big]$ $\quad A_1, A_2, A_3$ - random

$$\partial F(x, A) = \begin{cases} 1 + A_1 + A_2, & \text{if } x_0 + x_1 A_1 + x_2 A_2 + A_3 > 0 \\ -1 - A_1 - A_2, & \text{if } x_0 + x_1 A_1 + x_2 A_2 + A_3 < 0 \end{cases}$$

$$x_0^{N+1} = x_0^1 + \sum_{k=1}^{N} (\pm \tau_k), \qquad \sum_{k=1}^{\infty} (\pm) \frac{1}{k} < +\infty \quad \text{if } \pm \text{ equally likely.}$$



$$f_\delta(x) = \int_{-\delta}^{\delta} f(x-v)\frac{1}{2\delta}\,dv$$

Uniform kernel