Preference Robust Optimization: Characterization and Numerical Methods

Wenjie Huang

School of Data Science, The Chinese University of Hong Kong, Shenzhen
Group for Research in Decision Analysis

joint work with William B. Haskell, Jian Wu (Purdue U.) and Huifu Xu (CUHK)

GERAD Seminar

June 2rd 2021

Overview

Motivation: Homeland Security Budget Allocation I

(Hu et al, 2011)

Budget allocation problem for $m = 10$ U.S. cities subject to possible terrorist attack:

- Three underlying loss scenarios: {reduced loss, standard loss, increased loss}
- Measure loss in terms of $n = 4$ attributes: {property loss, fatalities, air departures, bridge traffic}

Random loss in city i for attribute j is C_{ij}.

Motivation: Homeland Security Budget Allocation II

- Data on property loss and fatalities are taken from (Willis et al, 2006)
- Daily bridge traffic and airport departures are assumed to follow a log-uniform distribution $\mathbb{P}(U = -1) = \mathbb{P}(U = 0) = \mathbb{P}(U = 1) = \frac{1}{3}$.
  1. Random incidents then satisfy $T_{ij} = \kappa_{ij}\gamma^U$ for a constant κ_{ij} and $\gamma > 1$
  2. Constant κ_{ij} depends on average values
  3. The cost satisfies $C_{ij} := c_i T_{ij}$ where c_i is the economic loss per incident of attribute i

Table 1. Terrorism losses, air departure, and average daily bridge traffic

	Property losses ($ million)			*Fatalities*				
Urban area	*Standard*	*Reduced*	*Increased*	*Standard*	*Reduced*	*Increased*	*Air departures*	*Average daily bridge traffic*
New York	413	265	550	304	221	401	23 599	596 400
Chicago	115	77	150	54	38	73	39 949	318 800
San Francisco	57	38	81	24	16	36	19 142	277 700
Washington, DC-MD-VA-WV	36	21	59	29	16	48	17 253	254 975
Los Angeles-Long Beach	34	16	58	17	7	31	28 816	336 000
Philadelphia, PA-NJ	21	8	28	9	5	13	13 640	192 204
Boston, MA-NH	18	8.3	26	12	8	17	11 625	669 000
Houston	11	6.7	15	9	6	12	20 979	308 060
Newark	7.3	0.8	12	4	0.1	9	12 827	518 100
Seattle-Bellevue-Everett	6.7	4	10	4	3	6	13 578	212 000

Motivation: Homeland Security Budget Allocation III

- Let z_{ij} be the amount invested in city i for attribute j:
  1. Set of feasible decisions $\mathcal{Z} := \{z \in \mathbb{R}^m : \sum_{i=1}^m \sum_{j=1}^n z_{ij} \leq B\}$ for budget e.g., $B = \$400$ million
  2. Investment functions $g_{ij}(z_{ij}) = v_{ij}(1 - \exp(-\delta z_{ij}))$ for $\delta \in (0, 1]$ (Nikoofal 2012)

- Given $z \in \mathcal{Z}$, define shortfall

$$C_i(z) := \sum_{j=1}^{m} \max\{C_{ij} - g_{ij}(z_{ij}), 0\}$$

for attribute $i = 1, ..., n$. Define vector $C(z) := (C_i(z))_{i=1}^n$ of shortfalls.

- How to find the best allocation plan $z^* \in \mathcal{Z}$?

Motivation: Homeland Security Budget Allocation IV

Compare two allocation plans: expected shortfall minimization vs. equal allocation

Urban area	*Property losses ($ million)*	*Fatalities ($ million)*	*Air departures ($ million)*	*Average daily bridge traffic ($ million)*
New York	62.472	32.253	1.000	9.985
Chicago	8.017	3.485	1.000	5.314
San Francisco	4.520	1.631	1.000	5.194
Washington	3.648	2.841	1.000	5.484
Los Angeles	4.284	1.973	1.000	9.194
Philadelphia	2.196	1.000	1.000	5.825
Boston	2.658	1.502	1.000	59.961
Houston	1.888	1.374	1.000	22.488
Newark	2.401	1.631	1.000	59.961
Seattle	4.456	2.401	1.000	59.961

Attribute	Loss *($ million)*		
	Scenario I	Scenario II	Scenario III
Property losses	0	0	79.57
Fatalities	0	0	17.74
Air Departures	6.87	9.02	12.58
Average daily bridge traffic	55.94	86.55	129.30

Attribute	Loss *($ million)*		
	Scenario I	Scenario II	Scenario III
Property losses	216.27	68.27	353.27
Fatalities	107.27	24.27	204.27
Air Departures	0	0	0
Average daily bridge traffic	267.95	314.80	365.95

Motivation: Why are these decisions difficult to make?

Patients Screening under the COVID-19 Pandemic (Roseli et al, 2020), Portfolio Optimization, Capital Allocation (Esfahani & Kuhn 2018)...

1. Something important is at stake and a person/group is held accountable for the decision.
2. The performance measure has multiple dimensions/attributes/criteria.
3. The alternatives are numerous.
4. Numerical optimization can only help once the decision maker's subjective preferences have been fully characterized.

Motivation: Von Neumann-Morgenstern Expected Utility

(Von Neumann & Morgenstern, 2007)

If the decision maker (DM) agrees with the following axioms:

1. Completeness : He can order any two lotteries.
2. Transitivity: $X_1 \succeq X_2 \succeq X_3 \Rightarrow X_1 \succeq X_3$
3. Continuity: If $X_1 \succeq X_2 \succeq X_3$ then there is a p such that $X_2 \sim pX_1 + (1-p)X_3$
4. Independence: If $X_1 \succeq X_2$, then $pX_1 + (1-p)X_3 \succeq pX_2 + (1-p)X_3$ for all p and X_3,

then there exists a function u such that:

$$X_1 \succeq X_2,$$

if and only if

$$\mathbb{E}[u(X_1)] \geq \mathbb{E}[u(X_2)].$$

Motivation: The Limitations of Utility Theory

- One can easily provide false information about his preferences. (Grable & Lytton, 1999)

  You are on a TV game show and can choose one of the following. Which would you take?

  A. $1,000 in cash
  B. A 50% chance at winning $5000
  C. A 25% chance at winning $10,000
  D. A 5% chance at winning $100,000

- One cannot specify a utility function in group decision-making where there must be a consensus.

Motivation: The Limitations of Utility Theory

- One may not agree all VNM axioms (Allais, 1953).

Experiment 1				Experiment 2			
Gamble 1A		Gamble 1B		Gamble 2A		Gamble 2B	
Winnings	Chance	Winnings	Chance	Winnings	Chance	Winnings	Chance
\$1 million	89%	\$1 million	89%	Nothing	89%	Nothing	89%
\$1 million	11%	Nothing	1%	\$1 million	11%	Nothing	1%
		\$5 million	10%			\$5 million	10%

$$\mathbb{E}[u(X)] = p_1 \cdot u(x_1) + p_2 \cdot u(x_2) + \ldots$$

This is a counterexample to the independence axiom. An explanation is the "certainty effect" (Tversky & Kahneman, 1986).

- One cannot make choice without full knowledge of the probability distribution.

Motivation: What is the right structure for the preferences ?

- Define and formulate the choice and preferences in new ways;
- Make simplifying assumptions about the structure of preference in order to allow interpolation and filter the errors;
- Employ a scheme that handles uncertainty about the preference;
- Develop scalable algorithms for decision-making with preferences;

Problem Formulation I

- A probability space $(\Omega, \mathcal{F}, \mathbb{P})$;
- Let $\mathcal{L} = \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^n)$ with $n \geq 1$ denote the set of essentially bounded random variables $X : \Omega \to \mathbb{R}^n$.
- The essential supremum norm

$$\|X\|_{\mathcal{L}} := \inf\{a \in \mathbb{R} : \mathbb{P}\{\|X(\omega)\|_\infty > a\} = 0\},$$

  where $\|\cdot\|_\infty$ is the ∞-norm on $\mathbb{R}^n$;
- Write $X = (X_j)_{j=1}^n$, where X_j represents attribute $j = 1, ..., n$.

Problem Formulation II

- A partial order $\succeq$ on $\mathcal{L}$ is a preference relation (i.e., weak order, total pre-order) if $\succeq$ is complete and transitive:
  (i) $\succeq$ is *complete* if for any $X, Y \in \mathcal{L}$, either $X \succeq Y$ or $Y \succeq X$ holds.
  (ii) $\succeq$ is *transitive* if $X \succeq Y$ and $Y \succeq Z$ holds, then $X \succeq Z$.
- Let $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$. The function $\phi : \mathcal{L} \to \bar{\mathbb{R}}$ is a *choice function* corresponding to "$\succeq$", then for any $X, Y \in \mathcal{L}$, we have $X \succeq Y$, if and only if $\phi(X) \geq \phi(Y)$.

Global Information:

1. [Mon](Monotone) $X \geq Y$ implies $\phi(X) \geq \phi(Y)$.
2. [QCo](Quasi-concave) For any $\lambda \in [0, 1]$,
$$\phi(\lambda X + (1-\lambda)Y) \geq \min\{\phi(X), \phi(Y)\}.$$
3. [Usc](Upper semi-continuous) $\limsup_{Y \to X} \phi(Y) = \phi(X)$.
4. [Lip](Lipschitz continuity) There exists $L > 0$, such that
$$|\phi(X) - \phi(Y)| \leq L \|X - Y\|_{\mathcal{L}}.$$

Problem Formulation III

Local Information:

- [Eli] (Preference elicitation) For a sequence of pairs of prospects $\mathcal{E} = \{(W_k, Y_k)\}_{k=1}^K$ exposed to the DM, the DM prefers W_k to Y_k for all $k = 1, \ldots, K$. Call $\mathcal{E}$ the elicited comparison data set (ECDS).

- [Nor] (Normalization) $\phi(W_0) = 0$ for some fixed normalizing prospect $W_0 \in \mathcal{L}$.

Problem Formulation IV

- Let $\mathscr{R}_{QCo}$ denote the set of all choice functions satisfying *Global Information*.
- Let $\mathscr{R}(\mathcal{E}) \subset \mathscr{R}_{QCo}$ be the *preference ambiguity set* satisfying *Local Information*.

Definition 1 (Robust choice function)

LThe *robust choice function* $\psi_{\mathscr{R}(\mathcal{E})} : \mathcal{L} \to \mathbb{R}$ corresponding to $\mathscr{R}(\mathcal{E})$ is defined by:

$$\psi_{\mathscr{R}(\mathcal{E})}(X) := \inf_{\phi \in \mathscr{R}(\mathcal{E})} \phi(X), \forall X \in \mathcal{L}.$$

Given a benchmark $Y \in \mathcal{L}$, the *robust choice function* with benchmark is defined via: $\psi_{\mathscr{R}(\mathcal{E})}(X; Y) := \inf_{\phi \in \mathscr{R}(\mathcal{E})}\{\phi(X) - \phi(Y)\}, \forall X \in \mathcal{L}$.

- **Why considering the "worst-case" paradigm?**

Problem Formulation V

- Let $\mathcal{Z}$ be a compact convex subset of a Euclidean space that represents feasible decisions. Let $G : \mathcal{Z} \to \mathcal{L}$ be a stochastic function that maps decisions in $\mathcal{Z}$ to prospects in $\mathcal{L}$ such that:

$$G(z, \omega) := [G(z)](\omega) \text{ is concave for all } \omega \in \Omega.$$

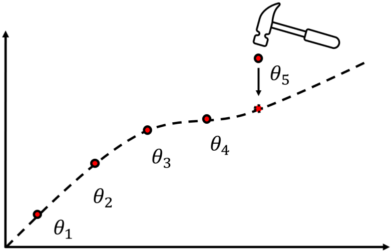
- We seek to maximize the stochastic function G w.r.t the robust choice function $\psi_{\mathscr{R}(\mathcal{E})}$. The preference robust optimization (PRO) is:

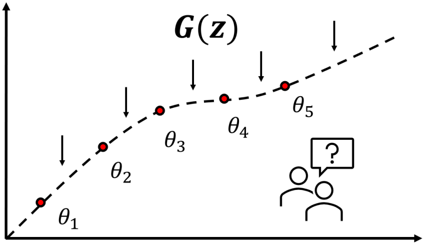
$$\max_{z \in \mathcal{Z}} \psi_{\mathscr{R}(\mathcal{E})}(G(z)) \equiv \max_{z \in \mathcal{Z}} \inf_{\phi \in \mathscr{R}(\mathcal{E})} \phi(G(z)).$$

1. Expected utility (Armbruster & Delage 2015, Hu & Mehrotra 2015, Haskell et al, 2016, Hu & Stepanyan 2017, Hu et al, 2018);
2. Risk measure (Delage & Li 2017, Delage et al, 2018, Wang & Xu 2020, Zhang et al, 2020, Guo & Xu 2021);
3. Target-based measure (Brown & Sim 2009, Brown et al, 2012);

The "Value" and "Interpolation" Problem I

- We start by studying the robust choice function, i.e., evaluating $\psi_{\mathscr{R}(\mathcal{E})}(G(z))$ for given $z \in \mathcal{Z}$.
- Define support set for local information $\Theta := \{W_0\} \cup \{(W_k,\ Y_k\}_{k=1}^{K}$ with total number of $J = 2K + 1$ prospects;
- Two-stage decomposition: the "value" problem and the "interpolation" problem;





The "Value" and "Interpolation" Problem II

- The "value" problem is:

$$\mathcal{P} := \inf_{\phi \in \mathscr{R}(\mathcal{E})} \sum_{\theta \in \Theta} \phi(\theta).$$

- The "interpolation" problem: Given any set of values $v = (v_\theta)_{\theta \in \Theta}$ for the choice function on the prospects in Θ,

$$\mathcal{P}(X;\, v) := \inf_{\phi \in \mathscr{R}_{QCo}} \{\phi(X) : [Lip],\, \phi(\theta) \geq v_\theta,\, \forall \theta \in \Theta\}.$$

Theorem 2

Problem $\mathcal{P}$ has a unique optimal solution $\phi^* \in \mathscr{R}_{QCo}$. Furthermore, for any $X \in \mathcal{L}$, $\psi_{\mathscr{R}(\mathcal{E})}(X) = val(\mathcal{P}(X;\, v^*))$, where $v^*_\theta = \phi^*(\theta) = \psi_{\mathscr{R}(\mathcal{E})}(\theta)$ for all $\theta \in \Theta$.

The “Value” and “Interpolation” Problem III

Assumption

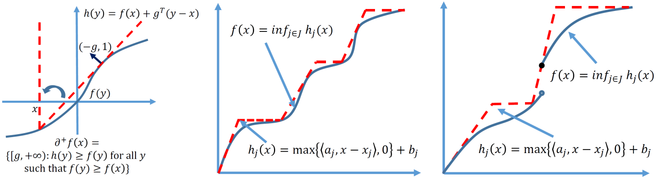
$\Omega = \{\omega_1, \omega_2, \ldots, \omega_T\}$ *(i.e., the underlying sample space is finite) and* $\mathbb{P}(\omega) > 0$ *for all* $\omega \in \Omega$ *(i.e., all scenarios have positive probability).*

- Identify a prospect $X \in \mathcal{L}$ with the vector of its realizations $\vec{X} = (X(\omega))_{\omega\in\Omega}$. equate $\vec{X} \equiv X$ and $\psi_{\mathscr{R}(\mathcal{E})}(\vec{X}) \equiv \psi_{\mathscr{R}(\mathcal{E})}(X)$ for all $X \in \mathcal{L}$. (Delage & Li 2017).

The "Value" and "Interpolation" Problem IV

- The disjunctive programming reformulation: under finite sample space assumption, and define $\widehat{\mathcal{E}}$ to be the set of all edges, the value problem $\mathcal{P}$ is equivalent to

$$
\begin{aligned}
\mathcal{P} \equiv \min_{v,\, s} \quad & \sum_{\theta \in \Theta} v_\theta && \text{(1a)} \\
\text{s.t.} \quad & v_\theta + \max\left\{\langle s_\theta,\, \theta' - \theta\rangle,\, 0\right\} \geq v_{\theta'}, \quad \forall\,(\theta,\, \theta') \in \mathcal{E}, && \text{(1b)} \\
& s_\theta \geq 0,\ \|s_\theta\|_1 \leq L, \quad \forall \theta \in \Theta, && \text{(1c)} \\
& v_\theta \geq v_{\theta'}, \quad \forall\,(\theta,\, \theta') \in \widehat{\mathcal{E}}, && \text{(1d)} \\
& v_{W_0} = 0. && \text{(1e)}
\end{aligned}
$$



The "Value" and "Interpolation" Problem V

- The interpolation problem $\mathcal{P}(\vec{X};\, v)$ is then equivalent to the disjunctive programming problem:

$$\mathcal{P}(\vec{X};\, v) \equiv \min_{a,\, b} \quad b \tag{2a}$$

$$\text{s.t.} \quad \max\left\{\langle a,\, \theta - \vec{X}\rangle,\, 0\right\} + b \geq v_\theta, \forall \theta \in \Theta, \tag{2b}$$

$$a \geq 0,\, \|a\|_1 \leq L. \tag{2c}$$

- Problem (1) and (2) can be turned into the mixed-integer linear program (MILP) reformulation (e.g., Big-M or convex hull).
- **One has to introduce J^2 binary variables to Problem (1) and J binary variables to Problem (2) !**

A Sorting Algorithm I

- Define an order list $\mathcal{D} := \{(\theta,\ v^*_\theta)\}_{\theta\in\Theta'}$ and $\mathcal{D}_t$ to denote the first t elements of $\mathcal{D}$, for $t = 1, 2, ..., J$. ($\mathcal{D}_1 = \{(W_0,\ 0)\}$ and $\mathcal{D}_J = \mathcal{D}$).
- We want to predict the value of $\psi_{\mathscr{R}(\mathcal{E})}(\theta)$, $\theta \notin \mathcal{D}_t$ by the following LP:

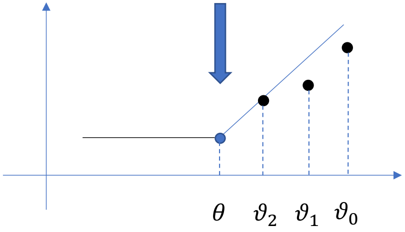
$$
\begin{aligned}
\mathscr{P}(\theta;\ \mathcal{D}_t) := \min_{v_\theta, s_\theta} \quad & v_\theta && \text{(3a)}\\
\text{s.t.} \quad & v_\theta + \left\langle s_\theta, \theta' - \theta \right\rangle \geq v^*_{\theta'},\ \forall \theta' \in \mathcal{D}_t, && \text{(3b)}\\
& s_\theta \geq 0, \|s_\theta\|_1 \leq L,\ \forall \theta \in \Theta, && \text{(3c)}\\
& v_\theta \geq v^*_{\theta'}, \quad \forall (\theta, \theta') \in \widehat{\mathcal{E}}, \theta' \in \mathcal{D}_t. && \text{(3d)}
\end{aligned}
$$

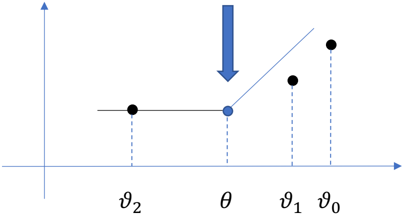
A Sorting Algorithm II

- Define the predictor of $\psi_{\mathscr{R}(\mathcal{E})}(\theta)$, $\theta \notin \mathcal{D}_t$ as

$$\pi(\theta; \mathcal{D}_t) := \min\left\{\underline{v}_t, \mathrm{val}(\mathscr{P}(\theta; \mathcal{D}_t))\right\},$$

where $\underline{v}_t := \min\{v^*_\theta \mid \theta \in \mathcal{D}_t\}$.





A Sorting Algorithm III

Algorithm 1: Sorting algorithm for the value problem

Initialization: Θ, $t = 1$, and $\mathcal{D}_t = \{(W_0, 0)\}$;
while $t < J$ **do**
  Choose $\theta^* \in \arg\max_{\theta \notin \mathcal{D}_t} \pi(\theta; \mathcal{D}_t)$, and set $u_{\theta^*} := \pi(\theta^*; \mathcal{D}_t)$;
  Set $\mathcal{D}_{t+1} := \{\mathcal{D}_t, (\theta^*, u_{\theta^*})\}$, and set $t := t + 1$;
end
return $\mathcal{D} := \mathcal{D}_J$.

Theorem 3

Algorithm 1 finds a decomposition $\mathcal{D}$ of Θ and computes $v^* = (v^*_\theta)_{\theta \in \Theta}$, after solving $O(J^2)$ linear programs.

- The MILP reformulation of Problem $\mathcal{P}$ may require $O(2^{J^2})$ LPs !

Law Invariance I

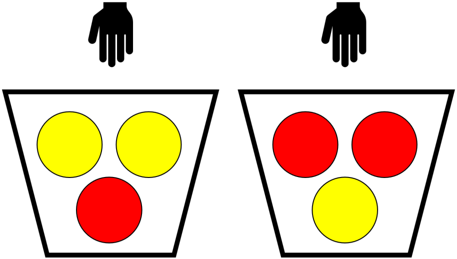


Figure: 2$ yellow ball and 1$ red ball (left); 2$ red ball and 1$ yellow ball (right)

- [Law] (Law invariance) $\phi(X) = \phi(Y)$ for all $X, Y \in \mathcal{L}$ such that $X =_D Y$.

Law Invariance II

- Assume that the probability measure $\mathbb{P}$ is uniform.
- The *permutations* on Ω is $\sigma(\Omega) = \{\omega_{\sigma(1)}, \omega_{\sigma(2)}, \ldots, \omega_{\sigma(T)}\}$. The permuted long vector is $\sigma(\vec{X}) = (X(\omega_{\sigma(t)}))_{t=1}^{T}$.
- When [Law] is in effect, $X \succeq Y$ is equivalent to $\phi(\sigma(\vec{X})) \geq \phi(\sigma'(\vec{Y}))$ for all $\sigma, \sigma' \in \Sigma$.
- The interpolation problem is

$$\min_{v_\theta, s_\theta} \quad v_\theta \tag{4a}$$

$$\text{s.t.} \quad v_\theta + \left\langle s_\theta, \sigma(\theta') - \theta \right\rangle \geq v^*_{\theta'}, \ \forall \theta' \in \mathcal{D}_{L,t},\ \sigma \in \Sigma, \tag{4b}$$

$$s_\theta \geq 0, \|s_\theta\|_1 \leq L, \ \forall \theta \in \Theta, \tag{4c}$$

$$v_\theta \geq v^*_{\theta'}, \quad \forall (\theta, \theta') \in \widehat{\mathcal{E}}, \theta' \in \mathcal{D}_{L,t}. \tag{4d}$$

Law Invariance III

- The reduce LP $\mathscr{P}_L(\theta; \mathcal{D}_{L,t})$ is

$$\min_{s, v_\theta, \{y_\theta, w_\theta\}_{\theta \in \mathcal{D}_{L,t}}} \quad v_\theta \tag{5a}$$

$$\text{s.t.} \quad \vec{1}^\top y_{\theta'} + \vec{1}^\top w_{\theta'} - \langle s, \theta \rangle + v_\theta - v^*_{\theta'} \geq 0, \quad \forall \theta' \in \mathcal{D}_{L,t}, \tag{5b}$$

$$\sum_{i=1}^{n} \theta'_i s_i^\top - y_{\theta'} \vec{1}^\top - \vec{1} w_{\theta'}^\top \geq 0, \quad \forall \theta' \in \mathcal{D}_{L,t}, \tag{5c}$$

$$s \geq 0, \|s\|_1 \leq L, \tag{5d}$$

$$v_\theta \geq v^*_{\theta'}, \quad \forall (\theta, \theta') \in \widehat{\mathcal{E}}, \theta' \in \mathcal{D}_{L,t}. \tag{5e}$$

- Based on Problem $\mathscr{P}_L(\theta; \mathcal{D}_{L,t})$, we can also develop a sorting algorithm for law invariance case.

Law Invariance IV

- An alternative representation: Can we define the binary relation and choice functions directly on the space of cumulative distribution functions ? **Yes !**

- Let P_X be the push-forward probability measure on $(\mathbb{R}^n, \mathcal{B})$ induced by $X \in \mathcal{L}$, defined by $P_X(B) := \mathbb{P}(X^{-1}(B))$ for all $B \in \mathcal{B}$. A lower orthant at $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ is

$$B = \{b \in \mathbb{R}^n : \ b_1 \leq x_1, \ b_2 \leq x_2, \ldots, \ b_n \leq x_n\}.$$

- Write $F_X(x)$ for $\mathbb{P}(X^{-1}(B))$. Define the set of CDFs associated with $\mathcal{L}$ to be

$$\mathscr{F}(\mathbb{R}^n) := \{F_X : X \in \mathcal{L}\}.$$

Law Invariance V

Let $F_1, F_2 \in \mathscr{F}(\mathbb{R}^n)$, then F_1 is said to be preferred to F_2 in the lower orthant order, written $F_1 \succeq_{lo} F_2$, if $F_1(x) \leq F_2(x)$ for all $x \in \mathbb{R}^n$.

A function $\phi : \mathscr{F}(\mathbb{R}^n) \to \bar{\mathbb{R}}$ is called a *choice function* on CDFs if it satisfies the following property:

- [Mon] (Monotonicity) For all $F_1, F_2 \in \mathscr{F}(\mathbb{R}^n)$, $F_1 \succeq_{lo} F_2$ implies $\phi(F_1) \geq \phi(F_2)$.

The set $\mathcal{X} = \{x^1, x^2, \dots, x^d\} \subset \mathbb{R}^n$ is finite and lexicographically ordered, then the CDFs F is step-like with breakpoints in $\mathcal{X}$ and

$$\vec{F} = (F(x^1), F(x^2), \dots, F(x^d)) \in \mathbb{R}^d.$$

1. Acceptability functional (Frittelli et al, 2014);
2. Yaari's dual theory of choice (Yaari, 1987);

Optimization Algorithms

1. **Binary Search:** Developed based on acceptance set representation;
2. **Level Search Method (LSM):** Developed based on the "interpolation" problem and a trick of getting ride of the disjunctive term;

- All of them rely on the decomposition into the "value" and "interpolation" problem and sorting algorithm for accelerate computation of "value" problem.

Optimization Algorithms: Binary Search I

- Normalizing by $\psi_{\mathscr{R}(\mathcal{E})}(\vec{W}_0) = 0$, the acceptance sets of $\psi_{\mathscr{R}(\mathcal{E})}$ is denoted as: $\mathcal{A}_v := \{\vec{X} \in \mathcal{L} : \psi_{\mathscr{R}(\mathcal{E})}(\vec{X}) \geq v\}, \forall v \leq 0$.
- **Main Idea:** With the acceptance sets of $\psi_{\mathscr{R}(\mathcal{E})}$ in hand, we can solve Problem (PRO) by doing binary search over the levels of the acceptance sets. Given level $v \leq 0$, we want to find some $z \in \mathcal{Z}$ such that $\vec{G}(z) \in \mathcal{A}_v$. If we can find such a z, then we can next search at a higher level; otherwise, we next search at a lower level.

Optimization Algorithms: Binary Search II

Theorem 4

Choose level $v \leq 0$ *and* $t = \kappa(v) := \{t = 1, 2, \ldots, J+1 \mid \underline{v}_{t+1} < v \leq \underline{v}_t\}$*. Then, there exists* $z \in \mathcal{Z}$ *such that* $\vec{G}(z) \in \mathcal{A}_v$ *if and only if* $\mathscr{F}_v(\mathcal{D}_t)$ *has a solution, where*

$$\mathscr{F}_v(\mathcal{D}_t) := \left\{ (z,\, p) \mid \vec{G}(z) \geq \sum_{\theta \in \mathcal{D}_t} \tilde{\theta} \cdot p_\theta + v/L,\ \sum_{\theta \in \mathcal{D}_t} p_\theta = 1,\ z \in \mathcal{Z},\ p \geq 0 \right\},$$

and $\tilde{\theta} := \theta - v_\theta^*/L$*.*

Theorem 5

Choose level $v \leq 0$ *and* $t = \kappa(v)$*, then* $\max_{z \in \mathcal{Z}} \psi_{\mathscr{R}(\mathcal{E})}(\vec{G}(z)) \geq v$ *if and only if* $val(\mathscr{G}(\mathcal{D}_t)) \geq v$*. where*

$$\mathscr{G}(\mathcal{D}_t) := \max_{z,p,v} \left\{ v : \mathscr{F}_v(\mathcal{D}_t) \text{ is feasible} \right\}.$$

Optimization Algorithms: Binary Search III

Algorithm 2: Binary search for Problem (PRO)

Initialization: $h_1 = H$, $h_2 = 0$;
while $h_1 \neq h_2 + 1$ **do**
  Set $h := \lceil \frac{h_1+h_2}{2} \rceil$, $t := \kappa(v^*_{[h-1]})$, and compute $v_t = \text{val}(\mathscr{G}(\mathcal{D}_t))$ with optimal solution z^*;
  if $v_t > \underline{v}_{t+1}$ **then** set $h_1 := h$;
  else set $h_2 := h$;
end
Set $h := \lceil \frac{h_1+h_2}{2} \rceil$, $t := \kappa(v^*_{[h-1]})$, and compute $v_t = \text{val}(\mathscr{G}(\mathcal{D}_t))$ with optimal solution z^*;
return z^* and $\psi_{\mathscr{R}(\mathcal{E})}(G(z^*)) = \min\{v_t, \underline{v}_t\}$.

Theorem 6

Algorithm 2 returns an optimal solution z^* of Problem (PRO), after solving $O(\log H)$ ($H \leq J$) instances of Problem $\mathscr{G}(\mathcal{D}_t)$.

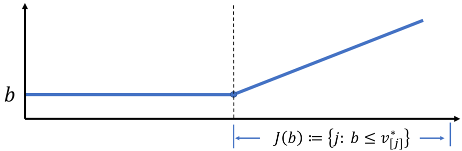
Optimization Algorithms: Level Search I

- Recall the interpolation problem:

$$\begin{aligned} \mathcal{P}(\vec{X};\, v^*) \equiv \min_{a,\, b} \quad & b \\ \text{s.t.} \quad & \max\left\{\langle a,\, \theta - \vec{X}\rangle,\, 0\right\} + b \geq v^*_\theta, \quad \forall \theta \in \Theta, \\ & a \geq 0,\ \|a\|_1 \leq L. \end{aligned}$$

- **Let's get ride of the disjunctive term !**

$$\max\left\{\langle a,\, \theta - \vec{X}\rangle,\, 0\right\} + b = \begin{cases} \langle a,\, \theta - \vec{X}\rangle + b & \langle a,\, \theta - \vec{X}\rangle \geq 0, \\ b & \langle a,\, \theta - \vec{X}\rangle < 0. \end{cases}$$



Optimization Algorithms: Level Search II

- Define $\mathcal{J}(b) := \{j \in \mathcal{J} : b \leq v_j^*\}$ to be the set $\theta_j \in \Theta$ for which $\psi_{\mathcal{R}(\mathcal{E})}(\theta_j) = v_j^* \geq b$. For any fixed level b, define:

$$\psi(\vec{X};\, b) := \min_{a \leq 0,\, \|a\|_\infty \leq L} \max_{j \in \mathcal{J}(b)} \left\{ v_j^* - \langle a,\, \theta_j - \vec{X} \rangle \right\}.$$

- For any fixed level b consider the following optimization problem:

$$\vartheta(b) := \max_{z \in \mathcal{Z}} \psi\left(\vec{G}(z);\, b\right).$$

Propose the univariate optimization problem for solving (PRO):

$$\min_{b \leq 0} \{b : b \geq \vartheta(b)\}. \tag{6}$$

- **Problem (6) and Problem (PRO) are equivalent !**

Optimization Algorithms: Level Search III

Algorithm 3: Level Search Method (LSM)

Step 1: Select initial range $[b_{\min}, 0]$, $b_l = b_{\min}$ and $b_u = 0$, set tolerance $\epsilon > 0$ and set $t = 0$.

Step 2: Let $b^t = (b_l + b_u)/2$. Check if $\vartheta(b^t) \geq b^t$, update $b_l \leftarrow b^t$; otherwise, update $b_u \leftarrow b^t$.

Step 3: If $b_u - b_l \leq \epsilon$, stop; otherwise, set $t := t + 1$; go to Step 2.

Theorem 7

Choose $\epsilon > 0$, *let* $\{b^t\}_{t \geq 0}$ *be produced by Algorithm 3, and let* $z^t \in \arg\max_{z \in \mathcal{Z}} \vartheta(b^t)$ *for all* $t \geq 0$. *Then, for all* $t \geq \log_2(|b_{\min}|/\epsilon)$, z^t *is an* $\epsilon-$*optimal solution of (PRO).*

Applications

- Guide a DM for decision-making. No one knows his choice function explicitly, including himself!
- Data-driven: Pairwise prospects are exposed to the DM, and then ECDS can be constructed.
- Decision is made in clairvoyant's perspective, such that we
  1. show the performance of the learned robust choice function.
  2. show the performance of the robust optimal solution in terms of the perceived value.

Applications

The true preference of DM is described by "perceived" choice function:

- *Perceived choice function (PCF) I*:

$$\phi_{CE}(X) = u^{-1}(\mathbb{E}[u(\langle w, X\rangle)]), \forall X \in \mathcal{L},$$

  where $w \in \mathbb{R}^n_+$ is a vector of weights. Take the piece-wise utility function

$$u(x) = \begin{cases} 1 - \exp(-\gamma x) & \text{if } x \geq 0, \\ \gamma x & \text{if } x < 0, \end{cases}$$

  with $\gamma = 0.05$.

- *Perceived choice function (PCF) II*:

$$\phi_{CE}(F_X) := u^{-1}\left(\int u(x) dF_X(x)\right).$$

Applications: Capital Allocation I

(Esfahani & Kuhn 2018)

- Consider a financial institution consisting of N sub-units, represented by the random vector $X = (X_1(\omega), X_2(\omega), \ldots, X_N(\omega))_{\omega \in \Omega}$.
- Define $\mathcal{Z} = \{Z \in \mathcal{L} : \sum_{n=1}^{N} Z_n(\omega) \leq B,\ Z(\omega) \geq 0,\ \forall \omega \in \Omega\}$, to be the set of admissible scenario-dependent financial recourse decisions subject to a budget constraint $B = 0.5$.
- A systematic risk factor $\varphi \sim Normal(0, 2\%)$ common to all sub-units. An idiosyncratic risk factor $\xi_n \sim Normal(n \times 3\%, n \times 2.5\%)$ specific to sub-unit $n = 1, 2, \ldots, N$. The return is set to be $X_n = 10(\varphi + \xi_n)$.
- Randomly generate samples of financial returns for ECDS construction. Pairwise preference is determined by using the PCF I.
- Our goal is to solve: $\max_{Z \in \mathcal{Z}} \psi_{\mathscr{R}(\mathcal{E})}(X + Z)$.
- Test sorting algorithm for value problem and binary search for PRO problem.

Applications: Capital Allocation II

Method	Group	ECDS Pairs 10	20	30	40	50
Sorting	1	15.6s	97s	364s	841s	1631s
	2	13.2s	107s	350s	826s	1539s
	3	12.5s	99s	326s	808s	1483s
	4	12.6s	107s	342s	823s	1651s
	5	13s	97s	337s	781s	1545s
	Average	**13.38s**	**101.4s**	**343.8s**	**815.8s**	**1569.8s**
MILP	1	7.16s	56.13s	185.78s	3728.40s	22109.66s
	2	4.64s	59.92s	254.80s	5312.21s	31019.20s
	3	6.47s	90.27s	450.12s	6560.72s	38534.88s
	4	5.30s	71.62s	387.61s	5666.33s	36112.33s
	5	5.18s	38.9s	278.50s	5449.45s	32334.57s
	Average	**5.75s**	**69.48s**	**296.90**	**5343.42s**	**32022.13s**

Table: Scalability of the sorting algorithm

Applications: Capital Allocation III

Settings	Group	ECDS Pairs						
		1	2	5	10	20	50	80
20 Scen, 20 Attr	1	52.9ms	437ms	7.19s	52.4s	506s	6715s	28284s
	2	47.5ms	420ms	7.12s	52.8s	507s	6822s	30111s
	3	52ms	445ms	7.28s	52.5s	508s	6720s	27765s
	4	50.2ms	440ms	7.2s	53s	511s	6669s	28202s
	5	51.8ms	444ms	7.22s	52.4s	512s	6733s	28116s
	Average	**50.88ms**	**437.2ms**	**7.202s**	**52.62s**	**508.8s**	**6731.8s**	**28495.6s**
Settings	Group	Number of Scenarios						
		5	10	15	20	50	100	200
10 Pairs, 20 Attr	1	6.64s	17s	33.4s	54.3s	274s	1141s	4532s
	2	6.75s	17.1s	33.2s	55.1s	274s	1222s	4605s
	3	6.58s	16.8s	33.3s	52.8s	280s	1088s	4520s
	4	6.32s	17.9s	33.6s	54.2s	269s	1050s	4330s
	5	6.67s	16.9s	33.8s	55s	271s	1150s	4567s
	Average	**6.592ms**	**17.14s**	**33.46s**	**54.28s**	**273.6s**	**1130.2s**	**4510.8s**
Settings	Group	Number of Attributes						
		5	10	15	20	50	100	200
10 Pairs, 20 Scen	1	21s	31s	40.9s	51.5s	124s	246s	480s
	2	20s	31s	40.6s	51.2s	122s	244s	484s
	3	19s	31s	40.7s	51.6s	123s	248s	482s
	4	20s	30s	41.2s	52.2s	126s	250s	480s
	5	21s	32s	41s	51.8s	124s	246s	481s
	Average	**20.2s**	**31s**	**40.88s**	**51.66s**	**123.8s**	**246.8s**	**481.4s**

Table: Scalability of the sorting algorithm (law-invariant case)

Applications: Capital Allocation IV

Settings	Group	ECDS Pairs 1	2	5	10	20	50	80
20 Scen, 20 Attr	1	576ms	867ms	1.21s	1.46s	1.84s	2.38s	4.56s
	2	532ms	902ms	1.17s	1.51s	1.82s	2.32s	4.62s
	3	555ms	855ms	1.15s	1.45s	1.81s	2.4s	4.88s
	4	546ms	832ms	1.22s	1.47s	1.85s	2.33s	4.67s
	5	612ms	841ms	1.23s	1.46s	1.9s	2.35s	4.51s
	Average	**564.2ms**	**859.4ms**	**1.196s**	**1.47s**	**1.844s**	**2.356s**	**4.648s**
		Number of Scenarios						
Settings	Group	5	10	15	20	50	100	200
10 Pairs, 20 Attr	1	153ms	414ms	1.02s	1.46s	8.8s	40s	181s
	2	115ms	414ms	924ms	1.36s	9.79s	35s	177s
	3	130ms	376ms	919ms	1.32s	10s	38s	190s
	4	119ms	427ms	917ms	1.33s	7.88s	42s	176s
	5	131ms	425ms	922ms	1.42s	9.47s	40s	183s
	Average	**129.6ms**	**417.2ms**	**940.4ms**	**1.378s**	**9.188s**	**39s**	**181.4s**
		Number of Attributes						
Settings	Group	5	10	15	20	50	100	200
10 Pairs, 20 Scen	1	402ms	745ms	1.42s	1.6s	4.56	9.42s	18.5s
	2	398ms	740ms	1.38s	1.55s	4.6s	9.44s	18.7s
	3	396ms	740ms	1.37s	1.56s	4.61s	9.45s	19s
	4	408ms	750ms	1.45s	1.61s	4.6s	9.39s	18.8s
	5	405ms	744ms	1.44s	1.61s	4.56s	9.4s	18.9s
	Average	**401.8ms**	**743.8ms**	**1.412s**	**1.586s**	**4.586s**	**9.42s**	**18.78s**

Table: Scalability of the binary search algorithm

Applications: Capital Allocation V

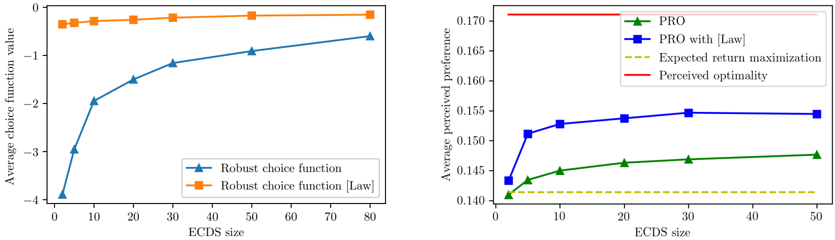


Figure: Robust choice function (left) and Performance of allocation (right) with ECDS size

Applications: Portfolio Optimization I

- The goal is to find the optimal mixture strategy of the m investment plans. The data consists of the daily return rates of exchange traded funds (ETFs) and the US central bank (FED) from January 2006 to December 2016, attained from Yahoo! Finance.
- Randomly choose a batch of A assets and then equally allocate wealth (normalized to one). The daily return rate of the investment plan X is the average of daily return rate of A assets.
- Let $y_{\min}$ and $y_{\max}$ denote lower and upper bounds on the daily returns of all assets. Approximate the interval $[y_{\min}, y_{\max}]$ with a uniform grid of J breakpoints satisfying $y_{\min} = y_1 \leq y_2 \leq \cdots \leq y_J = y_{\max}$. Let N be the total number of days. The empirical CDF is then:

$$\hat{F}(y_j) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}\{x_n \leq y_j\}, \, j = 1, \ldots, J.$$

Applications: Portfolio Optimization II

Some findings from the experiment:

- A pure strategy may outperform any mixture of investment plans.

  A randomized strategy will not outperform a deterministic one for choice functions satisfying the properties of translation invariance and 'mixture quasi-concavity' (Delage et al, 2019). We relax the condition because we do not enforce translation invariance.

Applications: Portfolio Optimization III

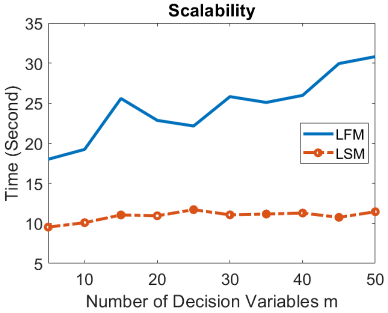
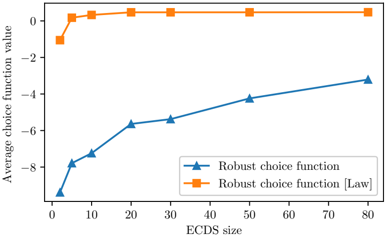


Figure: Scalability of LFM and LSM with number of investment plans m (left) and ECDS size (right)

Applications: Portfolio Optimization IV



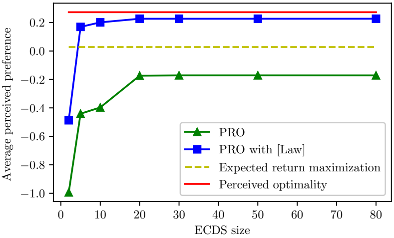


Figure: Robust choice function (left) and Performance of mixture strategy (right) with ECDS size

Conclusion

- We study preference robust optimization (PRO) problems where DM's preference is uncertain.
- Our choice function covers a number of well-known preference models.
- A data-driven scheme is employed to handle the uncertainty about the preference.
- Scalable algorithms are developed to compute the robust choice function and PRO problem.

Thank you !

Questions & Comments

References

- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, 503-546.
- Armbruster, B., Delage, E. (2015). Decision making under uncertainty when preference information is incomplete. *Management science*, 61(1), 111-128.
- Brown, D. B., Sim, M. (2009). Satisficing measures for analysis of risky positions. *Management Science*, 55(1), 71-84.
- Brown, D. B., Giorgi, E. D., Sim, M. (2012). Aspirational preferences and their representation by risk measures. *Management Science*, 58(11), 2095-2113.
- Delage, E., Guo, S., Xu, H. (2017). Shortfall risk models when information of loss function is incomplete. GERAD HEC Montréal.
- Delage, E., Li, J. Y. M. (2018). Minimizing risk exposure when the choice of a risk measure is ambiguous. *Management Science*, 64(1), 327-344.
- Delage, E., Kuhn, D., Wiesemann, W. (2019). "Dice"-sion–Making Under Uncertainty: When Can a Random Decision Reduce Risk?. *Management Science*, 65(7), 3282-3301.
- Esfahani, P. M., Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1), 115-166.

References

- Frittelli, M., Maggis, M., Peri, I. (2014). Risk measures on and value at risk with probability/loss function.*Mathematical Finance*, 24(3), 442-463.
- Grable, J., Lytton, R. H. (1999). Financial risk tolerance revisited: the development of a risk assessment instrument. *Financial services review*, 8(3), 163-181.
- Guo, S., Xu, H. (2021). Robust spectral risk optimization when the subjective risk aversion is ambiguous: a moment-type approach. *Mathematical Programming*, 1-36.
- Haskell, W. B., Fu, L., Dessouky, M. (2016). Ambiguity in risk preferences in robust stochastic optimization. *European Journal of Operational Research*, 254(1), 214-225.
- Hu, J., Homem-de-Mello, T., Mehrotra, S. (2011). Risk-adjusted budget allocation models with application in homeland security. *IIE Transactions*, 43(12), 819-839.
- Hu, J., Mehrotra, S. (2015). Robust decision making over a set of random targets or risk-averse utilities with an application to portfolio optimization. *IIE Transactions*, 47(4), 358-372.
- Hu, J., Stepanyan, G. (2017). Optimization with reference-based robust preference constraints. *SIAM Journal on Optimization*, 27(4), 2230-2257.
- Hu, J., Bansal, M., Mehrotra, S. (2018). Robust decision making using a general utility set. *European Journal of Operational Research*, 269(2), 699-714.

References

- Nikoofal, M. E., Zhuang, J. (2012). Robust allocation of a defensive budget considering an attacker's private information. *Risk Analysis: An International Journal*, 32(5), 930-943.
- Roselli, L. R. P., Frej, E. A., Ferreira, R. J. P., Alberti, A. R., de Almeida, A. T. (2020). Utility-based multicriteria model for screening patients under the COVID-19 pandemic. *Computational and Mathematical Methods in Medicine*, 2020.
- Tversky, A., Kahneman, D. (1986). Rational Choice and the Framing of Decisions. *The Journal of Business*, 59(4), S251-S278.
- Von Neumann, J., Morgenstern, O. (2007). *Theory of games and economic behavior* (commemorative edition). Princeton university press.
- Wang, W., Xu, H. (2020). Robust spectral risk optimization when information on risk spectrum is incomplete. *SIAM Journal on Optimization*, 30(4), 3198-3229.
- Willis, H. H., Morral, A. R., Kelly, T. K., Medby, J. J. (2006). *Estimating terrorism risk*. Rand Corporation.
- Xu, H. (2001). Level function method for quasiconvex programming. *Journal of Optimization Theory and Applications*, 108(2), 407-437.
- Yaari, M. (1987). The Dual Theory of Choice under Risk. *Econometrica*, 55(1), 95-115.
- Zhang, Y., Xu, H., Wang, W. (2020). Preference robust models in multivariate utility-based shortfall risk minimization. *Optimization Methods and Software*, 1-41.

Appendix: Example I

Example 8 (Robust utility maximization)

(Armbruster & Delage 2015, Hu & Mehrotra 2015, Haskell et al, 2016, Hu & Stepanyan 2017, Hu et al, 2018)

$$\max_{z\in\mathcal{Z}} \inf_{u\in\mathcal{U}} \mathbb{E}[u(G(z))].$$

$$\max_{z\in\mathcal{Z}} \inf_{u\in\mathcal{U}} u^{-1}(\mathbb{E}[u(G(z))]).$$

$$\max_{z\in\mathcal{Z}} \{f(z) : \text{s.t. } \mathbb{E}[u(G(z))] \geq \mathbb{E}[u(Y)], \forall u \in \mathcal{U}\}.$$

Example 9 (Robust risk exposure minimization)

(Delage & Li 2017, Delage et al, 2018, Wang & Xu 2020, Zhang et al, 2020, Guo & Xu 2021)

$$\min_{z\in\mathcal{Z}} \sup_{\rho\in\mathfrak{R}} \rho(-G(z)),$$

where ρ is monetary risk measure (convex, coherent, law invariant), e.g., spectral risk measure, utility-based shortfall...

Appendix: Example II

Example 10 (Target-based measure)

(Brown & Sim 2009, Brown et al, 2012): Given a family of risk measures ρ_k and target $\tau(k)$:

$$\mu(X) = \sup\{k \in \mathbb{R} : \rho_k(X - \tau(k)) \leq 0\}.$$

Conversely,

$$\tau(k) = \inf\{a \in \mathbb{R} : \mu(a) \geq k\},$$

$$\rho_k(X) = \inf\{a \in \mathbb{R} : \mu(X + a) \geq k\} - \tau(k).$$

1. Quasi-concave;
2. Connect to prospect theory;
3. Resolve many paradoxes: Allais, Ellsberg and Gain-loss separability;

Appendix: Example III

Example 11 (Acceptability functional for $n \geq 1$)

(Frittelli et al, 2014) Let $\{F_m\}_{m\in\mathbb{R}} \subset \mathscr{F}(\mathbb{R}^n)$ be a family of CDFs, and suppose $F_m(x)$ is decreasing in m for all fixed $x \in \mathbb{R}^n$. The corresponding acceptance sets are $\mathcal{A}_m := \{F_X \in \mathscr{F}(\mathbb{R}^n) : F_X(x) \leq F_m(x), \forall x \in \mathbb{R}^n\}$, the acceptability functional is

$$\phi(F_X) := \sup\{m \in \mathbb{R} : F_X(x) \leq F_m(x), \forall x \in \mathbb{R}^n\}.$$

Function ϕ is monotone increasing with respect to $\succeq_{lo}$ and quasi-concave.

Appendix: Example IV

Example 12 (Yaari's dual theory of choice)

(Yaari, 1987) Consider functional $\phi : \mathscr{F}(\mathbb{R}) \to \mathbb{R}$ defined as follows:

$$\phi(F_X) = \int_0^\infty g(1 - F_X(x))dx + \int_{-\infty}^0 [g(1 - F_X(x)) - 1]dx,$$

where $g : [0,1] \to [0,1]$ is a *strictly* increasing function with $g(0) = 0$ and $g(1) = 1$. Then ϕ is monotonically increasing w.r.t. $\succeq_{lo}$ and hence quasi-concave (also quasi-convex).

- Counterexample:
  Take $g(u) = \min\{\gamma\, u, 1\}$ for $\gamma > 1$. Choose CDFs $F_1(x) = \mathbb{I}_{x \geq 1}$ and $F_2(x) = \mathbb{I}_{x \geq 2}$. Let $\lambda \in [0,1]$ and $F_\lambda = \lambda F_1 + (1-\lambda)F_2$ be the compound distribution. Then $F_\lambda \succeq_{lo} F_2$ but $\phi(F_2) = \phi(F_\lambda) = 2$ for all $\lambda \in [1/\gamma, 1]$. The underlying reason is that the function $g(u) = 1$ is constant in the region $[1/\gamma, 1]$.