Inverse Stochastic Dominance Constraints Duality and Methods

Darinka Dentcheva[1] Andrzej Ruszczyński[2]

[1]Stevens Institute of Technology Hoboken, New Jersey, USA

[2]Rutgers University Piscataway, New Jersey, USA

Research supported by NSF awards DMS-0603728 and DMS-0604060

Outline

Stochastic Dominance

Distribution Functions

$$F_1(X;\eta) = \int_{-\infty}^{\eta} P_X(dt) = P\{X \le \eta\} \text{ for all } \eta \in \mathbb{R}$$

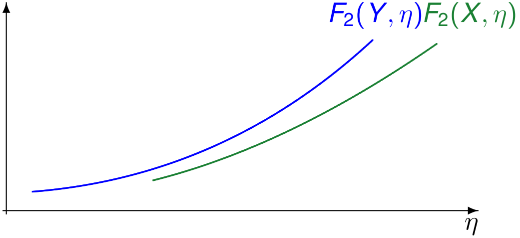
$$F_k(X;\eta) = \int_{-\infty}^{\eta} F_{k-1}(X;t)\,dt \text{ for all } \eta \in \mathbb{R}, \quad k = 2, 3, \ldots$$

kth order Stochastic Dominance

$$X \succeq_{(k)} Y \quad \Leftrightarrow \quad F_k(X,\eta) \le F_k(Y,\eta) \quad \text{for all } \eta \in \mathbb{R}$$

Second-Order Stochastic Dominance

$$F_2(X,\eta) = \int_{-\infty}^{\eta} F_1(X,t)\,dt = \mathbb{E}(\eta - X)_+ \text{ for all } \eta \in \mathbb{R}$$



The Lorenz Function (O. Lorenz, 1905)

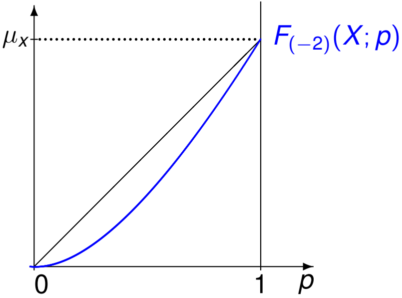
Quantile function

$$F_{(-1)}(X;p) = \inf\{\eta : F_1(X;\eta) \geq p\}$$

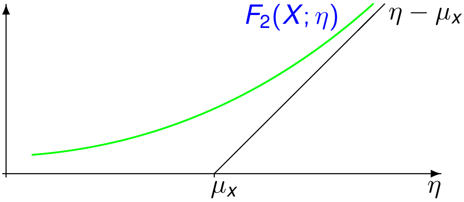
Absolute Lorenz function

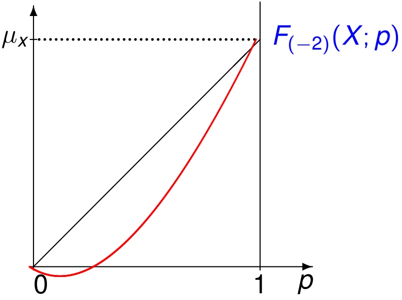
$$F_{(-2)}(X;p) = \int_0^p F_{(-1)}(X;t)\,dt \quad \text{for} \quad 0 < p \leq 1,$$

$F_{(-2)}(X;0) = 0$ and $F_{(-2)}(X;p) = +\infty$ for $p \notin [0,1]$.



Integrated Distribution Function and the Lorenz function



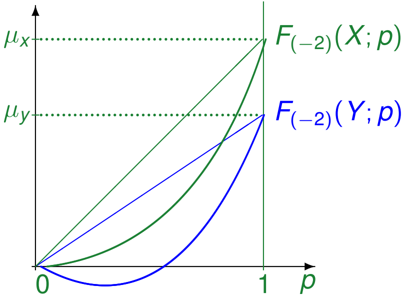


Fenchel conjugate function $\quad F^*(p) = \sup_u \{pu - F(u)\}.$

Ogryczak - Ruszczyński (2002)

$$F_{(-2)}(X;\cdot) = [F_2(X;\cdot)]^* \quad \text{and} \quad F_2(X;\cdot) = [F_{(-2)}(X;\cdot)]^*$$

Characterization of Stochastic Dominance by Lorenz Functions



$$X \succeq_{(2)} Y \quad \Leftrightarrow \quad F_{(-2)}(X;p) \geq F_{(-2)}(Y;p) \quad \text{for all} \quad 0 \leq p \leq 1.$$

Dominance Constrained Optimization

Given $Y \in \mathscr{L}_1(\Omega, \mathscr{F}, P)$ - benchmark random outcome

Inverse Stochastic Dominance Constraints

$$(\mathfrak{P}_{-2}) \qquad \begin{aligned} &\max f(z) \\ &\text{s.t. } F_{(-2)}(G(z);p) \geq F_{(-2)}(Y;p), \quad \forall\, p \in [\alpha, \beta], \\ &\qquad z \in Z \end{aligned}$$

Z is a closed subset of a vector space $\mathscr{Z}$, $[\alpha, \beta] \subset (0,1)$
$G : \mathscr{Z} \to \mathscr{L}_1(\Omega, \mathscr{F}, P)$ and $f : \mathscr{Z} \to \mathbb{R}$ are continuous.

Equivalent formulation

$$\max_{z, \eta(\cdot)} f(z)$$

$$\text{s.t. } \eta(p) - \frac{1}{p}\mathbb{E}\max(0, \eta(p) - G(z)) \geq \frac{1}{p}F_{(-2)}(Y;p) \quad \forall\, p \in [\alpha, \beta]$$

$$\eta(\cdot) - \text{nondecreasing}, \;\; z \in Z$$

with $\eta(p)$ a p-quantile of $G(z)$.

Discrete Equiprobable Probability Space

Assumption

$\Omega = \{\omega_1, \dots, \omega_N\}$ with probabilities $p_i = \mathbb{P}\{\omega_i\} = 1/N$, $i = 1, \dots, N$.

The Lorenz curves $F_{(-2)}(G(z); \cdot)$ and $F_{(-2)}(Y; \cdot)$ are convex and piecewise linear with break points at $\pi_i = i/N$.
For $\eta = (\eta_1, \dots, \eta_N)$, and $g_m(z) = [G(z)](\omega_m)$, set

$$\varrho_i = \frac{1}{\pi_i} F_{(-2)}(Y; \pi_i) = \frac{1}{i} \sum_{k=1}^{i} y_{[k]}.$$

Equivalent formulation

$$\begin{aligned} &\max_{z,\eta} f(z) \\ &\text{subject to } \eta_i - \frac{1}{i} \sum_{m=1}^{N} \max(0, \eta_i - g_m(z)) \geq \varrho_i \quad \forall\, i = 1, \dots, N \\ &z \in Z \end{aligned}$$

Conjugate Function Method

Step 0: Set $k = 0$.

Step 1: Solve the problem and obtain $(\hat{z}^k, \hat{\eta}^k)$:

$$\max_{z,\eta} f(z)$$

$$\text{s.t. } \eta_{i_j} - \frac{1}{i_j} \sum_{m \in A_j} (\eta_{i_j} - g_m(z)) \geq \varrho_{i_j} \quad \forall\, j = 1, \ldots, k$$

$$z \in Z$$

Step 2: Let $\delta_k = \max_{1 \leq i \leq N} \left\{ \varrho_i - \hat{\eta}_i^k + \frac{1}{i} \sum_{m=1}^{N} \max(0, \hat{\eta}_i^k - g_m(\hat{z}^k)) \right\}$.

If $\delta_k \leq 0$, stop; otherwise, continue.

Step 3: Let i_k be the index at which the maximum in Step 2 is achieved, and let $A_k = \{m : \hat{\eta}_{i_k}^k > g_m(\hat{z}^k)\}$.

Step 4: Increase k by one, and go to Step 1.

New Dominance Characterization for General Probability Spaces

Theorem

Suppose $X, Y \in \mathscr{L}_1(\Omega, \mathscr{F}, P)$. Then $X \succeq_{(2)} Y$ if and only if

$$\mathbb{E}[X|A] \geq \frac{1}{P(A)} F_{(-2)}(Y; P(A)) \quad \forall\, A \in \mathscr{F} :\ P(A) > 0.$$

Corollary $X \succeq_{(2)} Y$ if and only if

$$\mathbb{E}[X|X \leq t] \geq \frac{1}{F(X;t)} F_{(-2)}(Y; F(X;t)) \quad \forall\, t \in \mathbb{R},\ F(X;t) > 0.$$

Corollary Suppose X has a discrete distribution with N realizations, and $Y \in \mathscr{L}_1(\Omega, \mathscr{F}, P)$. Then $X \succeq_{(2)} Y$ if and only if

$$\sum_{k=1}^{j} \pi_k x_{[k]} \geq F_{(-2)}\Big(Y; \sum_{k=1}^{j} \pi_k\Big), \quad j = 1, \ldots, N.$$

Quantile Cutting Plane Methods: The Scaled Method

Step 0: Set $k = 0$.

Step 1: Solve the problem to obtain z^k and $X^k = G(z^k)$:

$$\max f(z)$$

$$\text{s.t. } \mathbb{E}[G(z)|S^j] \geq \frac{1}{P(S^j)} F_{(-2)}(Y; P(S^j)) \quad j = 1, \dots, k$$

$$z \in Z$$

Step 2: Consider the sets $A_t^k = \{X^k \leq t\}$ and let

$$\delta_k = \sup_t \left\{ \frac{1}{P(A_t^k)} F_{(-2)}(Y; P(A_t^k)) - \mathbb{E}[X^k | A_t^k] : P(A_t^k) > 0 \right\}.$$

If $\delta_k \leq 0$, stop; otherwise, continue.

Step 3: Find t_k such that $P(X^k \leq t_k) > 0$ and

$$\mathbb{E}[X^k | A_{t_k}^k] - \frac{1}{P(A_{t_k}^k)} F_{(-2)}(Y; P(A_{t_k}^k)) \leq -\frac{\delta_k}{2}.$$

Step 4: Set $S^{k+1} = A_{t_k}^k$, increase k by one, and go to Step 1.

Convergence of the Scaled Method

Theorem. Assume that $Z \subset \mathbb{R}^n$ is compact, $f(\cdot)$ is continuous, and the operator $G(\cdot) : \mathbb{R}^n \to \mathscr{L}_\infty(\Omega, \mathscr{F}, P)$ is Lipschitz continuous on Z. If problem $(\mathfrak{P}_{-2})$ has a nonempty feasible set, then either the scaled method stops at a solution of it, or every accumulation point of the sequence $\{z^k\}$ generated by the method is a solution of $(\mathfrak{P}_{-2})$.

In finite probability space Ω, the scaled method converges in finitely many iterations.

Modification

If the operator $G(\cdot) : \mathbb{R}^n \to \mathscr{L}_1(\Omega, \mathscr{F}, P)$ is Lipschitz continuous on Z, then we can modify the method to ensure convergence.

Quantile Cutting Plane Methods: The Modified Scaled Method

Assumption

The operator $G(\cdot) : \mathbb{R}^n \to \mathscr{L}_1(\Omega, \mathscr{F}, P)$ is Lipschitz continuous on Z.

We modify the method to ensure convergence by changing the following steps

Step 2a: Consider the sets $A_t^k = \{X^k \le t\}$ and let

$$\delta_k = \sup_t \left\{ \frac{1}{P(A_t^k)} F_{(-2)}(Y; P(A_t^k)) - \mathbb{E}[X^k | A_t^k] : P(A_t^k) \ge \varepsilon_k \right\}.$$

Step 2b: If $\delta_k \le 0$, replace ε_k by $\varepsilon_k/2$ and go to Step 2a; otherwise, continue.

Step 4: If $\delta_k < \varepsilon_k$ then set $\varepsilon_{k+1} = \min\{\delta_k, \varepsilon_k\}/2$; otherwise set $\varepsilon_{k+1} = \varepsilon_k$. Set $S^{k+1} = A_{t_k}^k$, increase k by one, and go to Step 1.

Observe that it is possible for the method to cycle between Steps 2a and 2b, without increasing the iteration index k.

Quantile Cutting Plane Methods: The Unscaled Method

The method uses cutting planes in a different form.

Step 2: Consider the sets $A_t^k = \{X^k = G(z^k) \le t\}$ and let

$$\delta_k = \sup_t \Big\{ F_{(-2)}(Y; P(A_t^k)) - F_{(-2)}(X^k; P(A_t^k)) : P(A_t^k) > 0 \Big\}.$$

If $\delta_k \le 0$, stop; otherwise, continue.

Step 3: Find t_k such that $P(X^k \le t_k) > 0$ as well as

$$F_{(-2)}(X^k; P(A_{t_k}^k)) - F_{(-2)}(Y; P(A_{t_k}^k)) \le -\frac{\delta_k}{2}.$$

Difference: no normalization by $P(A_t^k)$.

Theorem. Assume that $Z \subset \mathbb{R}^n$ is compact, $f(\cdot)$ is continuous, and the operator $G(\cdot) : \mathbb{R}^n \to \mathscr{L}_1(\Omega, \mathscr{F}, P)$ is Lipschitz continuous on Z. If problem $(\mathfrak{P}_{-2})$ has a nonempty feasible set, then either the unscaled method stops at a solution of $(\mathfrak{P}_{-2})$, or every accumulation point of the sequence $\{z^k\}$ is a solution of $(\mathfrak{P}_{-2})$.

Portfolio Optimization

Assets $j = 1, \dots, n$ with random return rates R_j
Reference return rate Y (e.g. index, existing portfolio, etc.)
Decision variables $z_j,\ j = 1, \dots, n$, Z -polyhedral set
Portfolio return rate $R(z) = \sum_{j=1}^{n} z_j R_j$

$$\begin{aligned} \max\ & f(z) \\ \text{s.t.}\ & \sum_{j=1}^{n} z_j R_j \succeq Y \\ & z \in Z \end{aligned}$$

$f(x) = \mathbb{E}\big[R(x)\big]$ or $f(x) = \varrho\big[R(x)\big]$: measure of risk.

Linear Programming Formulation

$$
\begin{aligned}
\text{maximize}\ & \sum_{i=1}^{N} p_i \sum_{k=1}^{n} r_{ik} z_k \\
\text{s.t.}\ & \sum_{k=1}^{n} r_{ik} z_k + s_{ij} \geq y_j, \quad i = 1, \dots, N,\ j = 1, \dots, N, \\
& \sum_{i=1}^{N} p_i s_{ij} \leq F_2(Y; y_j), \quad j = 1, \dots, N, \\
& s \geq 0, \quad z \in Z.
\end{aligned}
$$

Numerical Experience

Table: Dimensions of the three formulations.

$n = 500$ and Y is the return rate of the S&P 500 index.

Events	Direct LP		Conjugate Formulation		Quantile Formulations	
N	Variables	Constraints	Variables	Constraints	Variables	Constraints
50	3000	2551	550	51	500	1
100	10500	10101	600	101	500	1
150	23000	22651	650	151	500	1
200	40500	40201	700	201	500	1
250	63000	62751	750	251	500	1
300	90500	90301	800	301	500	1

Table: Performance on problems with equal probabilities of elementary events.

Events	Direct Linear Programming		Conjugate Function Method			Quantile Cutting Plane Methods Scaled			Unscaled		
N	CPU	Iter.	CPU	Cuts	Iter.	CPU	Cuts	Iter.	CPU	Cuts	Iter.
50	2.94	562	1.42	46	112	0.25	7	6	0.33	9	9
100	6.61	3170	7.54	147	697	0.80	26	61	1.00	32	73
150	21.23	9676	24.60	297	2664	1.09	36	182	1.56	50	261
200	393.20	15765	46.27	501	6430	1.61	48	135	1.79	56	149
250	905.11	34911	123.73	884	42206	2.19	62	477	2.42	67	608
300	-	-	174.19	1081	42790	3.34	89	637	3.99	105	562

Numerical Experience

Table: Performance on problems with unequal probabilities of elementary events.

Events	Direct Linear Programming		Quantile Cutting Plane Methods					
			Scaled			Unscaled		
N	CPU	Iter.	CPU	Cuts	Iter.	CPU	Cuts	Iter.
50	1.391	1044	0.718	23	98	0.749	24	59
100	7.953	4927	1.31	41	231	1.89	58	314
150	23.907	11548	1.313	41	219	2.309	68	510
200	144.171	21321	1.968	58	587	2.938	82	749
250	1928.98	34278	2.027	51	384	1.955	60	410
300	-	-	2.592	69	190	2.812	81	349

Orders and Utility Functions in Economics

For any two random variables $X, Y \in \mathscr{L}^1(\Omega, \mathscr{F}, P)$

Risk-Averse Consistency via Expected Utility

$X \succeq_{(2)} Y \Leftrightarrow \mathbb{E}\, u(X) \geq \mathbb{E}\, u(Y) \quad \forall$ nondecreasing concave $u : \mathbb{R} \to \mathbb{R}$.

Hadar and Russell (1969)

Risk-Averse Consistency via Rank Dependent Utility

$X \succeq_{(2)} Y$ holds true if and only if for all nondecreasing concave functions $w : [0, 1] \to \mathbb{R}$ that are subdifferentiable at 0

$$\int_0^1 F_{(-1)}(X; p)\, dw(p) \geq \int_0^1 F_{(-1)}(Y; p)\, dw(p).$$

Dentcheva and Ruszczyński (2006)

The Implied Rank Dependent Utility Function

Uniform inverse dominance condition (UIDC) for $(\mathfrak{P}_{-2})$

$\exists \tilde{z} \in Z$ such that $\inf_{p\in[\alpha,\beta]} \left\{ F_{(-2)}(G(\tilde{z}); p) - F_{(-2)}(Y; p) \right\} > 0.$

Lagrangian-like functional

$$\Phi(z, w) = f(z) + \int_0^1 F_{(-1)}(G(z); p)\, dw(p) - \int_0^1 F_{(-1)}(Y; p)\, dw(p)$$

$\mathscr{W}([\alpha, \beta])$ contains all concave and nondecreasing functions $w : [0, 1] \to \mathbb{R}$ such that $w(p) = 0$ for all $p \in [\beta, 1]$ and $w(p) = w(\alpha) + c(p - \alpha)$ with some $c > 0$, for all $p \in [0, \alpha]$.
The dual functional $\quad \Psi(w) = \sup_{z\in Z} \Phi(z, w).$

The dual problem

$(\mathfrak{D}_{-2}) \qquad \min_{w\in\mathscr{W}([\alpha,\beta])} \Psi(w).$

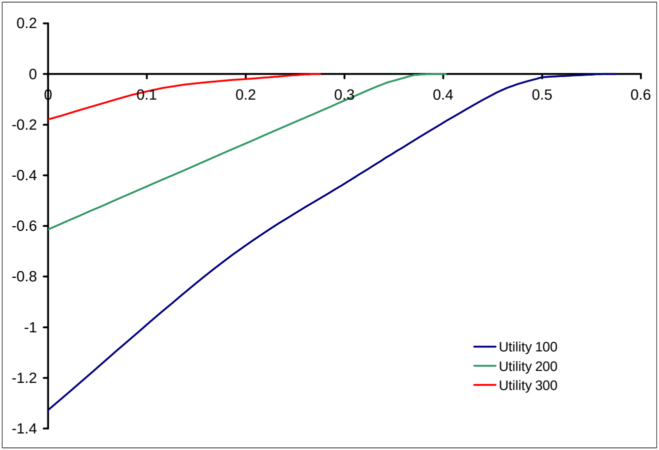
Theorem

Under the UIDC if problem $(\mathfrak{P}_{-2})$ has an optimal solution, then problem $(\mathfrak{D}_{-2})$ has an optimal solution and the optimal values of both problems coincide. The optimal solutions of the dual problem $(\mathfrak{D}_{-2})$ are the rank dependent utility functions $\hat{w} \in \mathscr{W}([\alpha, \beta])$ satisfying the dominance constraint and

$$\int_0^1 F_{(-1)}(G(\hat{z}); p)\, d\hat{w}(p) = \int_0^1 F_{(-1)}(Y; p)\, d\hat{w}(p)$$

for an optimal solution $\hat{z}$ of the problem $\max_{z \in Z} \Phi(z, \hat{w})$.

The Implied Rank Dependent Utility Function



A measure of risk ϱ assigns to an uncertain outcome $X \in \mathscr{L}_p(\Omega, \mathscr{F}, P)$ a real value $\varrho(X)$ on the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$.
A coherent measure of risk is a functional $\varrho : \mathscr{L}_1(\Omega, \mathscr{F}, P) \to \overline{\mathbb{R}}$ satisfying the axioms:

- Convexity: $\varrho(\alpha X + (1-\alpha)Y) \leq \alpha\varrho(X) + (1-\alpha)\varrho(Y)$for all $X, Y \in \mathscr{L}_1(\Omega, \mathscr{F}, P)$ and $\alpha \in [0,1]$.
- Monotonicity: If $X, Y \in \mathscr{L}_1(\Omega, \mathscr{F}, P)$ and $Y(\omega) \geq X(\omega)\ \forall \omega \in \Omega$, then $\varrho(Y) \leq \varrho(X)$.
- Translation Equivariance : If $a \in \mathbb{R}$ and $X \in \mathscr{L}_1(\Omega, \mathscr{F}, P)$, then $\varrho(X + a) = \varrho(X) - a$.
- Positive homogeneity: If $t > 0$ and $X \in \mathscr{L}_1(\Omega, \mathscr{F}, P)$, then $\varrho(tX) = t\varrho(X)$.

Artzner, Delbaen, Eber and Heath; Ruszczynski, Shapiro

Definition

Average Value at Risk of X at level p is defined as

$$\mathrm{AVaR}_p(X) = -\frac{1}{p} F_{(-2)}(X;p) = \frac{1}{p}\int_0^p \mathrm{VaR}_t(X)\,dt.$$

Theorem

For every law invariant, finite-valued coherent measure of risk on $\mathcal{L}_\infty(\Omega, \mathcal{F}, P)$ on atomless space Ω, a convex set $\mathscr{M} \subset \mathscr{P}((0,1])$ exists such that for all X

$$\varrho(X) = \sup_{\mu \in \mathscr{M}} \int_0^1 \mathrm{AVaR}_p(X)\,\mu(dp).$$

.

Definition

A measure of risk ϱ is called spectral if $\mathscr{M}$ is a singleton.

Mean-Risk Models as a Lagrangian Relaxation

Theorem (necessary optimality conditions):

Under the UIDC, if $\hat{z}$ is an optimal solution of $(\mathfrak{P}_{-2})$, then a spectral risk measure $\hat{\varrho}$ and a constant $\kappa \geq 0$ exist such that $G(\hat{z})$ is also an optimal solution of the problem

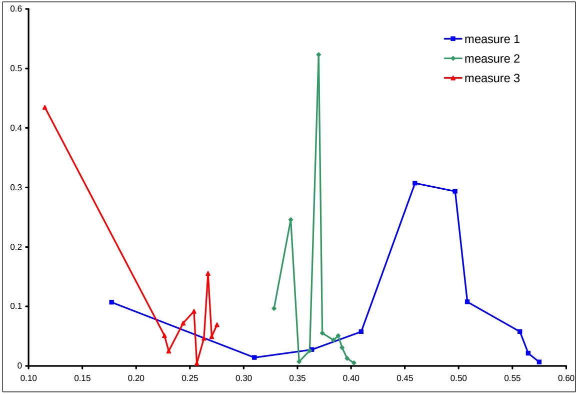
$$\max_{z \in Z} \left\{ f(z) - \kappa \hat{\varrho}(G(z)) \right\} \quad \text{and} \tag{1}$$

$$\kappa \hat{\varrho}(G(\hat{z})) = \kappa \hat{\varrho}(Y). \tag{2}$$

If the dominance constraint is active, then condition (2) takes on the form $\hat{\varrho}(G(\hat{z})) = \hat{\varrho}(Y)$.

The support of the measure μ in the spectral representation of $\hat{\varrho}(\cdot)$ is included in $[\alpha, \beta]$.

The Implied Measure of Risk



$$\begin{aligned}\varrho_{100}(X) &= 0.1069\mathrm{AVaR}_{0.1772}(X) + 0.014\mathrm{AVaR}_{0.3101}(X) + 0.0274\mathrm{AVaR}_{0.3636}(X)\\ &+ 0.0577\mathrm{AVaR}_{0.4093}(X) + 0.3073\mathrm{AVaR}_{0.4594}(X) + 0.2935\mathrm{AVaR}_{0.4967}(X)\\ &+ 0.1077\mathrm{AVaR}_{0.5081}(X) + 0.0576\mathrm{AVaR}_{0.557}(X) + 0.0213\mathrm{AVaR}_{0.5647}(X)\\ &+ 0.0066\mathrm{AVaR}_{0.575}(X)\end{aligned}$$