Electronic Companion to "Robust Optimization of Rank-Dependent Models with Uncertain Probabilities"

Guanyu Jin[a], Roger J. A. Laeven[a,c,d], and Dick den Hertog[b]

[a]Dept. of Quantitative Economics, University of Amsterdam, 1001 NJ Amsterdam, The Netherlands
[b]Dept. of Business Analytics, University of Amsterdam, 1001 NJ Amsterdam, The Netherlands
[c]EURANDOM, 5600 MB Eindhoven, The Netherlands
[d]CentER, Tilburg University, 5000 LE Tilburg, The Netherlands

April 14, 2025

Abstract

This text serves as an Electronic Companion to the paper "Robust Optimization of Rank-Dependent Models with Uncertain Probabilities". For context, notation and definitions, see the main paper. We summarize its contents:

- In EC.1, we provide Tables EC.1 and EC.2.
- In EC.2, we provide all the proofs.
- In EC.3, we provide detailed derivations of the conjugate functions and epigraphs for the collection of examples in Tables EC.1 and EC.2.
- In EC.4, we provide additional details on the reformulation of (11).
- In EC.5, we provide a visualization of the shapes of $\mathcal{U}_{\phi,h}(\mathbf{p})$.
- In EC.6, we provide further details on the optimistic dual counterpart.
- In EC.7, Algorithm EC.1, we provide the adaptation of Algorithm 1 to problem (P-constraint).
- In EC.8, we provide the SOS2 formulation of the nomimal problem (P-Nom-EG) with constraints (34), if the set $\mathcal{A}$ is polyhedral, $f(\mathbf{a},\mathbf{x}) = \mathbf{a}^T\mathbf{x}$ and u is piecewise-linear.
- In EC.9, we provide the details of the Hit-and-Run algorithm used in Section 7.

EC.1 Tables EC.1 and EC.2

Distortion family	*Function* $h(p),\ p \in [0,1]$	*Conjugate* $(-h)^*(y),\ y \leq 0$	*Epigraph of perspective* $\lambda(-h)^*\left(\frac{-\nu}{\lambda}\right) \leq z,\ \lambda > 0$	*Conic representation*
Expectation	p	$0,\ y \leq -1$	$0 \leq z,\ \nu \geq \lambda$	CQ
$\mathrm{CVaR}_{1-\alpha}$	$\min\{\frac{p}{1-\alpha}, 1\},\ \alpha \in [0,1)$	$\max\{(1-\alpha)y + 1, 0\}$	$\max\{-(1-\alpha)\nu + \lambda, 0\} \leq z,\ \nu \geq 0$	CQ
Proportional Hazard	$\begin{cases} p^r \\ r \in (0,1) \end{cases}$	$r^{\frac{1}{1-r}}(1+\frac{1}{r})\lvert y\rvert^{\frac{r}{r-1}}$	$\begin{cases} \lambda(r^{\frac{r}{1-r}} - r^{\frac{1}{1-r}})^{1-r} \leq z^{1-r}\nu^r \\ \nu \geq 0 \end{cases}$	PC
Absolute Deviation	$\begin{cases} (1+r)p & p < 1/2 \\ (1-r)p + r & p \geq 1/2 \\ r \in (0,1) \end{cases}$	$\max\{(y+1+r)/2, 0\},\ y \leq -(1-r)$	$\begin{cases} \max\{(-\nu + \lambda + \lambda r)/2, 0\} \leq z \\ -\nu \leq -(1-r)\lambda \end{cases}$	CQ
Gini Principles	$\begin{cases} (1+r)p - rp^2 \\ r \in (0,1) \end{cases}$	$\begin{cases} \frac{1}{4r}\max\{y+1+r, 0\}^2 & y+1 \leq r \\ y+1 & y+1 > r \end{cases}$	$\begin{cases} z = z_1 + z_2 \\ -\nu + \lambda(1-r) = \xi_1 + \xi_2 \\ (z_1 + \lambda) \geq \sqrt{\frac{w^2}{r} + (z_1 - \lambda)^2},\ z_2 \geq \xi_2 \\ \xi_1 + 2r \leq w \\ \xi_1 \leq 0, \xi_2, w, \nu \geq 0 \end{cases}$	CQ
Dual Moments	$\begin{cases} 1 - (1-p)^n \\ n > 1 \end{cases}$	$\begin{cases} y + c(n) \cdot \min\{\lvert y\rvert, n\}^{\frac{n}{n-1}} + 1 \\ c(n) = \left(n^{-\frac{1}{n-1}} - n^{-\frac{n}{n-1}}\right) \end{cases}$	$\begin{cases} \lambda - \xi_3 + (n^{-\frac{1}{n-1}} - n^{-\frac{n}{n-1}})\xi_4 \leq z \\ \xi_2 \leq \xi_4^{\frac{n-1}{n}} \cdot \lambda^{1-\frac{n-1}{n}} \\ -\nu + \xi_3 \leq 0,\ \xi_2 \geq \xi_3, \\ \xi_2, \xi_3, \xi_4 \geq 0. \end{cases}$	PC
MAXMINVAR	$\begin{cases} (1-(1-p)^n)^{1/n} \\ n > 1 \end{cases}$	$\begin{cases} y\left(1 - s^{\frac{1}{n}}\right) + (1-s)^{\frac{1}{n}} \\ s = \frac{\lvert y\rvert^{\frac{n}{n-1}}}{1+\lvert y\rvert^{\frac{n}{n-1}}} \end{cases}$	$\begin{cases} -\nu + \xi_3 \leq 0, \xi_2 \geq \lambda \\ -\xi_3 + \lVert(\xi_2, \xi_3)\rVert_{\frac{n}{n-1}} \leq z \\ \xi_2, \xi_3 \geq 0 \end{cases}$	PC
Lookback Transform	$\begin{cases} p^r(1 - \log(p^r)) \\ r \in (0,1) \end{cases}$	Closed form unknown	$\begin{cases} \xi_3 + \lambda e^{-(\xi_2+\xi_3)/\lambda} + (r^{\frac{r}{1-r}} - r^{\frac{1}{1-r}})\xi_4 \leq z \\ \xi_2 \leq \nu^r \xi_4^{1-r} \\ \xi_2, \xi_3, \xi_4, \nu \geq 0 \end{cases}$	EXP × PC

Table EC.1: Canonical distortion functions with explicit expressions of their conjugates $(-h)^*$ and conic representations of the epigraphs of their perspectives. The cones are abbreviated as: CQ: quadratic cone, PC: power cone, EXP: exponential cone. Note that $(-h)^*(y) = +\infty$ for $y > 0$; see Remark 1. For Expectation and $\mathrm{CVaR}_{1-\alpha}$, see Föllmer and Schied (2016), Section 4.6; for Proportional Hazard, see Wang (1995); for Absolute Deviation and Gini Principles, see Denneberg (1990b); for Dual Moments, see Muliere and Scarsini (1989) and Eeckhoudt et al. (2020); for MAXMINVAR, see Cherny and Madan (2009); for Lookback Transform, see Denneberg (1990a).

Divergence family	*Function* $\phi(x),\ x \geq 0$	*Conjugate* $\phi^*(y),\ y \in \mathbb{R}$	*Epigraph of perspective* $\gamma\phi^*\left(\frac{s}{\gamma}\right) \leq t,\ \gamma > 0$	*Conic representation*
Kullback-Leibler	$x \log x - x + 1$	$e^y - 1$	$\begin{cases} \gamma \log(\frac{\gamma}{w}) + s \leq 0 \\ w - \gamma \leq t. \end{cases}$	EXP
Burg entropy	$-\log x + x - 1$	$-\log(1-y),\ y < 1$	$\begin{cases} \gamma \log(\frac{\gamma}{v}) \leq t \\ v = \gamma - s,\ v > 0 \end{cases}$	EXP
χ^2-distance	$\frac{1}{x}(x-1)^2$	$2 - 2\sqrt{1-y},\ y < 1$	$\begin{cases} 2\gamma - 2w \leq t \\ \sqrt{w^2 + \frac{1}{4}(\gamma - v)^2} \leq \frac{1}{2}(\gamma + v) \\ w \geq 0,\ v = \gamma - s,\ v > 0 \end{cases}$	CQ
Variation distance	$\lvert x-1 \rvert$	$\max\{s, -1\},\ s \leq 1$	$\max\{s, -\gamma\} \leq t,\ s \leq \gamma$	CQ
Modified χ^2-distance	$(x-1)^2$	$\max\{0, y/2+1\}^2 - 1$	$\begin{cases} \sqrt{w^2 + \frac{t^2}{4}} \leq \frac{t+2\gamma}{2} \\ 0 \leq w, s/2 + \gamma \leq w \end{cases}$	CQ
Hellinger distance	$(\sqrt{x} - 1)^2$	$\frac{y}{1-y},\ y < 1$	$\begin{cases} -\gamma + v \leq t \\ \sqrt{\gamma^2 + \frac{1}{4}(v - w)^2} \leq \frac{1}{2}(v + w) \\ w = \gamma - s,\ w > 0. \end{cases}$	CQ
χ-divergence of order $\theta > 1$	$\lvert x-1 \rvert^\theta$	$y + (\theta - 1)(\frac{\lvert y \rvert}{\theta})^{\frac{\theta}{\theta-1}}$	$\begin{cases} s + (\theta-1)\theta^{\frac{\theta}{1-\theta}} w \leq t, 0 \leq w \\ \lvert s \rvert \leq w^{\frac{\theta}{\theta-1}} \cdot \gamma^{1-\frac{\theta}{\theta-1}} \end{cases}$	PC
Cressie and Read	$\frac{1-\theta+\theta x - x^\theta}{\theta(1-\theta)},\ 0 < \theta < 1$	$\begin{cases} \frac{1}{\theta}(1 - y(1-\theta))^{\frac{\theta}{\theta-1}} - \frac{1}{\theta} \\ y > \frac{1}{1-\theta} \end{cases}$	$\begin{cases} \lvert w \rvert \leq (t\theta + \gamma)^{\frac{\theta-1}{\theta}} \cdot \gamma^{1-\frac{\theta-1}{\theta}} \\ w = \gamma - s(1-\theta),\ s < \frac{\gamma}{1-\theta} \end{cases}$	PC

Table EC.2: Canonical ϕ-divergence functions taken from Table 2 of Ben-Tal et al. (2013), but with the explicit epigraphs of their perspective functions (provided in the fourth column) and corresponding conic representations (provided in the fifth column). The cones are abbreviated as: CQ: quadratic cone, PC: power cone, EXP: exponential cone.

EC.2 Proofs

Proof of Theorem 1. Let $\mathbf{q}$ be an arbitrary probability vector and denote by $\mathbb{Q}$ the corresponding probability measure on the sigma-algebra $2^{|\Omega|}$, i.e., the set of all subsets of Ω. Recall Eqn. (3). By Denneberg (1994), pp. 16–17, the set function induced by the composition $h \circ \mathbb{Q}$ is monotone and submodular. It then follows from Proposition 10.3 of Denneberg (1994) that, for any $X : \Omega \to \mathbb{R}$,

$$\rho_{u,h,\mathbf{q}}(X) = \sup_{\bar{\mathbf{q}} \in M_h(\mathbf{q})} \mathbb{E}_{\bar{\mathbf{q}}}[-u(X)],$$

with $M_h(\mathbf{q})$ defined in (15).

Hence, we have that

$$\begin{aligned}\sup_{\mathbf{q} \in \mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) &= \sup_{\mathbf{q} \in \mathcal{D}_\phi(\mathbf{p},r)} \sup_{\bar{\mathbf{q}} \in M_h(\mathbf{q})} \mathbb{E}_{\bar{\mathbf{q}}}[-u(f(\mathbf{a},\mathbf{X}))] \\ &= \sup_{(\mathbf{q},\bar{\mathbf{q}}) \in \mathcal{U}_{\phi,h}(\mathbf{p})} -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a},\mathbf{x}_i)),\end{aligned}$$

where $\mathcal{U}_{\phi,h}(\mathbf{p})$ is as defined in (13). This proves the stated result. $\square$

Proof of Theorem 2. Suppose that a pair $(\mathbf{a}, c) \in \mathbb{R}^{n_a+1}$ satisfies the inequality

$$\sup_{(\mathbf{q},\bar{\mathbf{q}}) \in \mathcal{U}_{\phi,h}(\mathbf{p})} -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a},\mathbf{x}_i)) \leq c. \tag{EC.1}$$

Then, the left-hand side of (EC.1) constitutes a maximization problem upper bounded by $c < \infty$. Moreover, the set $\mathcal{U}_{\phi,h}(\mathbf{p})$ contains $(\mathbf{p},\mathbf{p})$ as a Slater point, since $\phi(1) = 0 < r$ and $h(\sum_{k \in I_j} p_k) > \sum_{k \in I_j} p_k$ for all subsets $I_j \subset [m]$ that are not $\emptyset$ and $[m]$.[1] Therefore, strong duality holds, and we consider the Lagrangian function

$$\begin{aligned}&L(\mathbf{a},\mathbf{q},\bar{\mathbf{q}},\alpha,\beta,(\lambda_j)_j,\gamma) \\ &= -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a},\mathbf{x}_i)) - \alpha\left(\sum_{i=1}^m q_i - 1\right) - \beta\left(\sum_{i=1}^m \bar{q}_i - 1\right) - \sum_{j=1}^{2^m-2} \lambda_j \left(\sum_{k \in I_j} \bar{q}_k - h\left(\sum_{k \in I_j} q_k\right)\right) \\ &\quad - \gamma\left(\sum_{i=1}^m p_i \phi\left(\frac{q_i}{p_i}\right) - r\right) \\ &= \alpha + \beta + \gamma r + \sum_{i=1}^m -(u(f(\mathbf{a},\mathbf{x}_i)) + \beta)\bar{q}_i - \sum_{j=1}^{2^m-2} \lambda_j \sum_{k \in I_j} \bar{q}_k + \sum_{i=1}^m -\alpha q_i - \gamma\left(\sum_{i=1}^m p_i \phi\left(\frac{q_i}{p_i}\right)\right) \\ &\quad - \sum_{j=1}^{2^m-2} -\lambda_j h\left(\sum_{k \in I_j} q_k\right),\end{aligned}$$

for $\alpha, \beta \in \mathbb{R}$ and $\lambda_j, \gamma \geq 0$. We analyze $\sup_{\mathbf{q},\bar{\mathbf{q}} \geq 0} L(\mathbf{a},\mathbf{q},\bar{\mathbf{q}},\alpha,\beta,(\lambda_j)_j,\gamma)$, which excluding the

[1] This is because $h(x) > x$ for all $x \in (0,1)$ due to concavity (excluding the trivial case where $h(x) = x$ for all $x \in [0,1]$).

constant $\alpha + \beta + \gamma r$ is equal to:

$$\begin{aligned}
&\sup_{\mathbf{q},\bar{\mathbf{q}}\geq 0} \sum_{i=1}^{m} -(u(f(\mathbf{a},\mathbf{x}_i)) + \beta)\bar{q}_i - \sum_{j=1}^{2^m-2} \lambda_j \sum_{k\in I_j} \bar{q}_k + \sum_{i=1}^{m} -\alpha q_i - \gamma p_i \phi\left(\frac{q_i}{p_i}\right) \\
&\quad - \sum_{j=1}^{2^m-2} \lambda_j(-h)\left(\sum_{k\in I_j} q_k\right) \\
&= \sup_{\bar{\mathbf{q}}\geq 0} \sum_{i=1}^{m} -(u(f(\mathbf{a},\mathbf{x}_i)) + \beta)\bar{q}_i - \sum_{j=1}^{2^m-2} \lambda_j \sum_{k\in I_j} \bar{q}_k + \sup_{\mathbf{q}\geq 0} \sum_{i=1}^{m} -\alpha q_i - \gamma p_i \phi\left(\frac{q_i}{p_i}\right) \\
&\quad - \sum_{j=1}^{2^m-2} \lambda_j(-h)\left(\sum_{k\in I_j} q_k\right).
\end{aligned}$$

We examine both supremum terms separately. The first supremum gives:

$$\begin{aligned}
\sup_{\bar{\mathbf{q}}\geq 0} \sum_{i=1}^{m} -(u(f(\mathbf{a},\mathbf{x}_i)) + \beta)\bar{q}_i - \sum_{j=1}^{2^m-2} \lambda_j \sum_{k\in I_j} \bar{q}_k &= \sup_{\bar{\mathbf{q}}\geq 0} \sum_{i=1}^{m} -\left(u(f(\mathbf{a},\mathbf{x}_i)) + \beta + \sum_{j:i\in I_j} \lambda_j\right)\bar{q}_i \\
&= \begin{cases} 0 & \text{if } u(f(\mathbf{a},\mathbf{x}_i)) + \beta + \sum_{j:i\in I_j} \lambda_j \geq 0,\ \forall i \\ \infty & \text{else.} \end{cases}
\end{aligned}$$

The second supremum gives:

$$\begin{aligned}
&\sup_{\mathbf{q}\geq 0} \sum_{i=1}^{m} -\alpha q_i - \gamma p_i \phi\left(\frac{q_i}{p_i}\right) - \sum_{j=1}^{2^m-2} \lambda_j(-h)\left(\sum_{k\in I_j} q_k\right) \\
&= \sup_{\substack{\mathbf{q}\geq \mathbf{0} \\ w_1,\ldots,w_{2^m-2}\geq 0}} \left\{ \sum_{i=1}^{m} -\alpha q_i - \gamma p_i \phi\left(\frac{q_i}{p_i}\right) - \sum_{j=1}^{2^m-2} \lambda_j(-h)(w_j) \;\middle|\; w_j = \sum_{k\in I_j} q_k \right\} \\
&= \inf_{\nu_1,\ldots,\nu_{2^m-2}} \sup_{\substack{\mathbf{q}\geq \mathbf{0} \\ w_1,\ldots,w_{2^m-2}\geq 0}} \sum_{i=1}^{m} -\alpha q_i - \gamma p_i \phi\left(\frac{q_i}{p_i}\right) - \sum_{j=1}^{2^m-2} \lambda_j(-h)(w_j) - \sum_{j=1}^{2^m-2} \nu_j\left(w_j - \sum_{k\in I_j} q_k\right) \\
&= \inf_{\nu_1,\ldots,\nu_{2^m-2}} \sup_{\mathbf{q}\geq 0} \sum_{i=1}^{m} \left(-\alpha + \sum_{j:i\in I_j} \nu_j\right) q_i - \gamma p_i \phi\left(\frac{q_i}{p_i}\right) + \sup_{w_1,\ldots,w_{2^m-2}\geq 0} \sum_{j=1}^{2^m-2} -\nu_j w_j - \lambda_j(-h)(w_j) \\
&= \inf_{\nu_1,\ldots,\nu_{2^m-2}} \sum_{i=1}^{m} p_i \left(\sup_{t\geq 0} \left(-\alpha + \sum_{j:i\in I_j} \nu_j\right) t - \gamma\phi(t)\right) + \sum_{j=1}^{2^m-2} \lambda_j(-h)^*\left(\frac{-\nu_j}{\lambda_j}\right) \\
&= \inf_{\nu_1,\ldots,\nu_{2^m-2}} \sum_{i=1}^{m} p_i \gamma \phi^*\left(\frac{-\alpha + \sum_{j:i\in I_j} \nu_j}{\gamma}\right) + \sum_{j=1}^{2^m-2} \lambda_j(-h)^*\left(\frac{-\nu_j}{\lambda_j}\right),
\end{aligned}$$

with ϕ^* and $(-h)^*$ the conjugates of ϕ and $-h$.

Therefore, strong duality implies that a pair $(\mathbf{a}, c)$ satisfies (EC.1) if and only if

$$\inf_{\substack{\alpha,\beta,\nu_j\in\mathbb{R}\\ \lambda_j,\gamma\geq 0}} \alpha+\beta+\gamma r+\sum_{i=1}^{m} p_i\gamma\phi^*\left(\frac{-\alpha+\sum_{j:i\in I_j}\nu_j}{\gamma}\right)+\sum_{j=1}^{2^m-2}\lambda_j(-h)^*\left(\frac{-\nu_j}{\lambda_j}\right)\leq c$$
$$\text{subject to } -u(f(\mathbf{a},\mathbf{x}_i))-\beta-\sum_{j:i\in I_j}\lambda_j\leq 0,\ \forall i\in[m].$$

Since the infimum is attained due to the boundedness of the primal problem (EC.1) and the strong duality theorem, we may remove the infimum sign and obtain that the above holds if and only if there exist $\lambda_j,\gamma\geq 0,\alpha,\beta,\nu_j\in\mathbb{R}$ such that

$$\begin{cases} \alpha+\beta+\gamma r+\sum_{i=1}^{m} p_i\gamma\phi^*\left(\frac{-\alpha+\sum_{j:i\in I_j}\nu_j}{\gamma}\right)+\sum_{j=1}^{2^m-2}\lambda_j(-h)^*\left(\frac{-\nu_j}{\lambda_j}\right)\leq c\\ -u(f(\mathbf{a},\mathbf{x}_i))-\beta-\sum_{j:i\in I_j}\lambda_j\leq 0,\ \forall i\in[m]\\ \lambda_j,\gamma\geq 0\\ \alpha,\beta,\nu_j\in\mathbb{R},\ \forall j\in[2^m-2]. \end{cases}$$

For the nominal problem, the Lagrangian function is given by

$$\begin{aligned} L(\mathbf{a},\bar{\mathbf{q}},\beta,(\lambda_j)_j) &= -\sum_{i=1}^{m}\bar{q}_i u(f(\mathbf{a},\mathbf{x}_i))-\beta\left(\sum_{i=1}^{m}\bar{q}_i-1\right)-\sum_{j=1}^{2^m-2}\lambda_j\left(\sum_{k\in I_j}\bar{q}_k-h\left(\sum_{k\in I_j}p_k\right)\right)\\ &=\beta+\sum_{j=1}^{2^m-2}\lambda_j h\left(\sum_{k\in I_j}p_k\right)+\sum_{j=1}^{m}\bar{q}_i\left(-u(f(\mathbf{a},\mathbf{x}_i))-\beta-\sum_{j:i\in I_j}\lambda_j\right). \end{aligned}$$

Hence, we have

$$\begin{aligned} &\sup_{\bar{\mathbf{q}}\in M_h(\mathbf{p})}\sum_{i=1}^{m}-\bar{q}_i u(f(\mathbf{a},\mathbf{x}_i))\\ &=\inf_{\beta\in\mathbb{R},\lambda_j\geq 0}\sup_{\bar{\mathbf{q}}\geq\mathbf{0}} L(\mathbf{a},\bar{\mathbf{q}},\beta,(\lambda_j)_j)\\ &=\inf_{\beta\in\mathbb{R},\lambda_j\geq 0}\left\{\beta+\sum_{j=1}^{2^m-2}\lambda_j h\left(\sum_{k\in I_j}p_k\right)\ \middle|\ -u(f(\mathbf{a},\mathbf{x}_i))-\beta-\sum_{j:i\in I_j}\lambda_j\leq 0,\ \forall i\in[m]\right\}, \end{aligned}$$

where the above strong duality holds since we are solving a linear programming problem. Therefore, we have that $\sup_{\bar{\mathbf{q}}\in M_h(\mathbf{p})}\sum_{i=1}^{m}-\bar{q}_i u(f(\mathbf{a},\mathbf{x}_i))\leq c$ if and only if there exist $\beta\in\mathbb{R},(\lambda_j)_j\geq 0$ such that

$$\begin{cases} \beta+\sum_{j=1}^{2^m-2}\lambda_j h\left(\sum_{k\in I_j}p_k\right)\leq c\\ -u(f(\mathbf{a},\mathbf{x}_i))-\beta-\sum_{j:i\in I_j}\lambda_j\leq 0,\ \forall i\in[m]\\ \beta\in\mathbb{R},\lambda_j\geq 0,\ \forall j\in[2^m-2]. \end{cases}$$

$\square$

Proof of Lemma 1. The proof follows Ben-Tal and Nemirovski (2019), except that we replace the quadratic cone with a general cone. We have

$$\mathrm{Epi}(f^*)=\{(\mathbf{y},s):\mathbf{y}^T\mathbf{x}-f(\mathbf{x})\leq s,\forall\mathbf{x}\}=\{(\mathbf{y},s):\mathbf{y}^T\mathbf{x}-t\leq s,\forall(\mathbf{x},t)\in\mathrm{Epi}(f)\}.$$

Therefore, $(\mathbf{y}, s) \in \mathrm{Epi}(f^*)$ if and only if

$$\min_{\mathbf{x},t,\mathbf{w}} \{-\mathbf{y}^T\mathbf{x} + t : \mathbf{A}\mathbf{x} + t\mathbf{v} + \mathbf{B}\mathbf{w} + \mathbf{b} \succeq_{\mathbf{K}} \mathbf{0}\} \geq -s.$$

This minimization problem is strictly feasible and bounded from below. Hence, the conic duality theorem (see Ben-Tal and Nemirovski, 2019) implies that it is equal to

$$\max_{\boldsymbol{\xi}} \{-\mathbf{b}^T\boldsymbol{\xi} : \mathbf{A}^T\boldsymbol{\xi} = -\mathbf{y}, \mathbf{B}^T\boldsymbol{\xi} = \mathbf{0}, \mathbf{v}^T\boldsymbol{\xi} = 1, \boldsymbol{\xi} \in \mathbf{K}^*\}.$$

Therefore,

$$\mathrm{Epi}(f^*) = \{(\mathbf{y}, s) : \exists\, \boldsymbol{\xi} \in \mathbf{K}^* : \mathbf{A}^T\boldsymbol{\xi} = -\mathbf{y}, \mathbf{B}^T\boldsymbol{\xi} = \mathbf{0}, \mathbf{v}^T\boldsymbol{\xi} = 1, s \geq \mathbf{b}^T\boldsymbol{\xi}\}.$$

□

Proof of Lemma 2. Following Ben-Tal and Nemirovski (2019), we have

$$\begin{aligned}
\mathrm{Epi}(\tilde{f}) = \left\{(\mathbf{x}, \lambda, t) : \lambda f\left(\frac{\mathbf{x}}{\lambda}\right) \leq t\right\} &= \left\{(\mathbf{x}, \lambda, t) : \left(\frac{\mathbf{x}}{\lambda}, \frac{t}{\lambda}\right) \in \mathrm{Epi}(f)\right\} \\
&= \{(\mathbf{x}, \lambda, t) : \exists\, \mathbf{w} \in \mathbb{R}^k : \mathbf{A}(\mathbf{x}/\lambda, \mathbf{w}, t/\lambda)^T - b \succeq_{\mathbf{K}} \mathbf{0}\} \\
&= \{(\mathbf{x}, \lambda, t) : \exists\, \tilde{\mathbf{w}} \in \mathbb{R}^k : \mathbf{A}(\mathbf{x}/\lambda, \tilde{\mathbf{w}}/\lambda, t/\lambda)^T - b \succeq_{\mathbf{K}} \mathbf{0}\} \\
&= \{(\mathbf{x}, \lambda, t) : \exists\, \tilde{\mathbf{w}} \in \mathbb{R}^k : \mathbf{A}(\mathbf{x}, \tilde{\mathbf{w}}, t)^T - \lambda b \succeq_{\mathbf{K}} \mathbf{0}\} \\
&= \{(\mathbf{x}, \lambda, t) : \exists\, \tilde{\mathbf{w}} \in \mathbb{R}^k : [\mathbf{A}, -b](\mathbf{x}, \tilde{\mathbf{w}}, t, \lambda)^T \succeq_{\mathbf{K}} \mathbf{0}\}.
\end{aligned}$$

□

Proof of Lemma 3. By definition, $\mathbf{q}^* \in \mathcal{D}_\phi(\mathbf{p}, r)$. Hence, we only need to show that $\bar{\mathbf{q}}^* \in M_h(\mathbf{q}^*)$. Using Lemma 4.98 of Föllmer and Schied (2016), we have that $\rho_{u,h,\mathbf{q}^*}(X) \geq \mathbb{E}_{\bar{\mathbf{q}}^*}[-X]$ for all random variables X (where $\rho_{u,h,\mathbf{q}^*}(X)$ is as defined in (4) with measure $\mathbb{Q} = \mathbf{q}^*$). In particular, this holds for all $X = -\mathbb{1}_A$, for any measurable set $A \subset \Omega$. Hence, $(\mathbf{q}^*, \bar{\mathbf{q}}^*) \in \mathcal{U}_{\phi,h}(\mathbf{p})$. □

Proof of Theorem 3. First, we have, for any $\mathbf{a}^1, \mathbf{a}^2 \in \mathcal{A}$,

$$\begin{aligned}
\left|\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}^2, \mathbf{x}_i)) - \sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}^1, \mathbf{x}_i))\right| &\leq \|\bar{\mathbf{q}}\|_2 \|(u(f(\mathbf{a}^2, \mathbf{x}_i)) - u(f(\mathbf{a}^1, \mathbf{x}_i)))_{i=1}^m\|_2 \\
&\leq \|\bar{\mathbf{q}}\|_1 \|(u(f(\mathbf{a}^2, \mathbf{x}_i)) - u(f(\mathbf{a}^1, \mathbf{x}_i)))_{i=1}^m\|_2 \\
&= \|(u(f(\mathbf{a}^2, \mathbf{x}_i)) - u(f(\mathbf{a}^1, \mathbf{x}_i)))_{i=1}^m\|_2,
\end{aligned} \tag{EC.2}$$

where the first inequality follows from Cauchy-Schwarz; and the second inequality follows from $\|\bar{\mathbf{q}}\|_2 \leq \|\bar{\mathbf{q}}\|_1$ for a probability vector $\bar{\mathbf{q}}$, since $\bar{q}_k^2 \leq \bar{q}_k$ for $|\bar{q}_k| \leq 1$.

Suppose now that the cutting-plane method has not terminated at the t-th iteration, i.e., the optimal solution and objective value $(\mathbf{a}_t, c_t)$ violate the ϵ_{tol}-feasibility condition at step 4 of Algorithm 1. Let $(\mathbf{q}_t^*, \bar{\mathbf{q}}_t^*)$ be the new worst-case scenarios that are added to $\mathcal{U}_t$ at step 5 of Algorithm 1. Then, by definition, we have

$$-\sum_{i=1}^m \bar{q}_{t,i}^* u(f(\mathbf{a}_t, \mathbf{x}_i)) - c_t > \epsilon_{\mathrm{tol}}. \tag{EC.3}$$

For any $s > t$, we also have that at the s-th iteration:

$$-\sum_{i=1}^{m} \bar{q}_{t,i}^{*} u(f(\mathbf{a}_s, \mathbf{x}_i)) - c_s \leq 0, \tag{EC.4}$$

since this is part of the constraint at the s-th iteration. Let $\tilde{c}$ be the optimal objective value of (P-ref). By Assumption 1, we have that $-\infty < \tilde{c} < \infty$. Furthermore, we have $\tilde{c} \geq c_i$ for any iteration i since the cutting-plane algorithm always yields a lower bound on (P-ref). Hence, we may assume that for t and s sufficiently large, the lower bound improvement is upper bounded, i.e., $c_s - c_t \leq \frac{1}{2}\epsilon_{\text{tol}}$, since otherwise the cutting-plane will yield a lower bound that exceeds $\tilde{c}$, after finitely many iterations. Therefore, it follows from (EC.3)–(EC.4) that

$$\left| \sum_{i=1}^{m} \bar{q}_{t,i}^{*} u(f(\mathbf{a}_t, \mathbf{x}_i)) - \sum_{i=1}^{m} \bar{q}_{t,i}^{*} u(f(\mathbf{a}_s, \mathbf{x}_i)) \right| > \epsilon_{\text{tol}} - (c_s - c_t) \geq \frac{1}{2}\epsilon_{\text{tol}},$$

which implies that

$$\|(u(f(\mathbf{a}_s, \mathbf{x}_i)) - u(f(\mathbf{a}_t, \mathbf{x}_i)))_{i=1}^{m}\|_2 > \frac{1}{2}\epsilon_{\text{tol}}. \tag{EC.5}$$

This shows that the minimum distance between any two outcomes of the cutting-plane method, when evaluated in utilities, is at least $\frac{1}{2}\epsilon_{\text{tol}}$.

The idea is now to show that if the cutting-plane method does not terminate, then there is a sequence of infinitely many cutting-plane solutions $\{\mathbf{a}_j\}_{j=1}^{\infty}$ for which the corresponding vector $(u(f(\mathbf{a}_j, \mathbf{x}_i)))_{i=1}^{m}$ remains in a bounded set. Since we know from above that for each of these solutions their utility values are a distance $\frac{1}{2}\epsilon_{\text{tol}}$ away from each other, we conclude that the cutting-plane method must terminate since a bounded set can not contain infinitely many disjoint balls with radius $\frac{1}{2}\epsilon_{\text{tol}}$, as argued in Mutapcic and Boyd (2009).

Therefore, we define

$$\mathcal{T} \triangleq \left\{ (u(f(\mathbf{a}, \mathbf{x}_i)))_{i=1}^{m} \;\middle|\; \mathbf{a} \in \mathcal{A},\; -\sum_{i=1}^{m} p_i u(f(\mathbf{a}, \mathbf{x}_i)) \leq \tilde{c} \right\}.$$

Then, for all iterations t, we have $(u(f(\mathbf{a}_t, \mathbf{x}_i)))_{i=1}^{m} \in \mathcal{T}$, since $\mathbf{p} \in \mathcal{U}_t$ for all $t \geq 1$, hence we have the inequality $-\sum_{i=1}^{m} p_i u(f(\mathbf{a}_t, \mathbf{x}_i)) \leq c_t \leq \tilde{c}$. We show that $\mathcal{T}$ is bounded in the Euclidean 2-norm $\|.\|_2$. Indeed, by assumption, $M \triangleq \sup_{\mathbf{a}\in\mathcal{A}, i\in[m]} u(f(\mathbf{a}, \mathbf{x}_i)) < \infty$. Hence, for all vectors $(u(f(\mathbf{a}, \mathbf{x}_i)))_{i=1}^{m} \in \mathcal{T}$, its individual entry is always bounded from above by M. It remains to show that all its entries are also bounded from below. Let $p_{\min} = \min_{i=1}^{m} p_i > 0$ (by Assumption 2) and $i_{\min}(\mathbf{a}) = \text{argmin}_i u(f(\mathbf{a}, \mathbf{x}_i)$. Then, we have for all $(u(f(\mathbf{a}, \mathbf{x}_i)))_{i=1}^{m} \in \mathcal{T}$,

$$-p_{i_{\min}(\mathbf{a})} u(f(\mathbf{a}, \mathbf{x}_{i_{\min}(\mathbf{a})})) - \sum_{i \neq i_{\min}(\mathbf{a})} p_i u(f(\mathbf{a}, \mathbf{x}_i)) \leq \tilde{c},$$

which implies

$$\begin{gathered} -u(f(\mathbf{a}, \mathbf{x}_{i_{\min}(\mathbf{a})})) \leq \frac{\tilde{c} + M}{p_{i_{\min}(\mathbf{a})}} \leq \frac{\tilde{c} + M}{p_{\min}} \\ \Leftrightarrow \\ u(f(\mathbf{a}, \mathbf{x}_{i_{\min}(\mathbf{a})})) \geq -\frac{\tilde{c} + M}{p_{\min}} > -\infty. \end{gathered}$$

Hence, $\mathcal{T}$ is bounded in the Euclidean 2-norm $\|.\|_2$. $\square$

Proof of Theorem 4. Before we proceed with the proof, we recall the notion of a KKT-vector. For a generic convex optimization problem

$$\min_{x \in \mathcal{X}} \{f_0(x) \mid f_i(x) \leq 0, \forall i \in [L]\},$$

with $\mathcal{X}$ a suitable convex subset and $f_0, \ldots, f_L$ convex functions, the KKT-vector corresponding to the constraints $f_i(x) \leq 0, i \in [L]$ is a vector $\lambda \in \mathbb{R}^L$ such that

$$\inf_{x \in \mathcal{X}} \left\{ f_0(x) + \sum_{i=1}^{L} \lambda_i f_i(x) \right\} = \min_{x \in \mathcal{X}} \{f_0(x) \mid f_i(x) \leq 0, \forall i \in [L]\}.$$

The existence of the KKT-vector is guaranteed when Slater's condition and the boundedness of the optimization problem are satisfied (see Theorem 28.2, Rockafellar, 1970).

The proof now consists of two parts. In the first part, we show that Algorithm EC.1 terminates after finitely many iterations, for any $\epsilon_{\text{tol}} > 0$. The second part shows the convergence as $\epsilon_{\text{tol}} \to 0$.

First part: We reexamine the proof of Theorem 3. First, we have, for any $\mathbf{a}^1, \mathbf{a}^2 \in \mathcal{A}$, the following inequality:

$$\left| \sum_{i=1}^{m} \bar{q}_i u(f(\mathbf{a}^2, \mathbf{x}_i)) - \sum_{i=1}^{m} \bar{q}_i u(f(\mathbf{a}^1, \mathbf{x}_i)) \right| \leq \|(u(f(\mathbf{a}^2, \mathbf{x}_i)) - u(f(\mathbf{a}^1, \mathbf{x}_i)))_{i=1}^m\|_2.$$

Let t be an iteration where Algorithm EC.1 has not yet terminated and let $(\mathbf{q}_t^*, \bar{\mathbf{q}}_t^*)$ be the new worst-case scenarios that are added to $\mathcal{U}_t$. Then, we have $-\sum_{i=1}^m \bar{q}_{t,i}^* u(f(\mathbf{a}_t, \mathbf{x}_i)) - c > \epsilon_{\text{tol}}$. For any $s > t$, we also have that at the s-th iteration $-\sum_{i=1}^m \bar{q}_{t,i}^* u(f(\mathbf{a}_s, \mathbf{x}_i)) - c \leq 0$. Hence,

$$\left| \sum_{i=1}^{m} \bar{q}_{t,i}^* u(f(\mathbf{a}_t, \mathbf{x}_i)) - \sum_{i=1}^{m} \bar{q}_{t,i}^* u(f(\mathbf{a}_s, \mathbf{x}_i)) \right| > \epsilon_{\text{tol}},$$

which implies that

$$\|(u(f(\mathbf{a}_s, \mathbf{x}_i)) - u(f(\mathbf{a}_t, \mathbf{x}_i)))_{i=1}^m\|_2 > \epsilon_{\text{tol}}.$$

We define the set

$$\mathcal{T} \triangleq \left\{ (u(f(\mathbf{a}, \mathbf{x}_i)))_{i=1}^m \;\middle|\; \mathbf{a} \in \mathcal{A}, \; -\sum_{i=1}^{m} p_i u(f(\mathbf{a}, \mathbf{x}_i)) \leq c \right\}.$$

Then, for all iterations t, we have $(u(f(\mathbf{a}_t, \mathbf{x}_i)))_{i=1}^m \in \mathcal{T}$, since $\mathbf{p} \in \mathcal{U}_t$ for all $t \geq 1$. It follows from the proof of Theorem 3 that $\mathcal{T}$ must be bounded in the Euclidean 2-norm $\|.\|_2$. Hence, Algorithm EC.1 must terminate after finitely many steps.

Second part: Let $g(\mathbf{a}_{\epsilon_{\text{tol}}})$ be the objective value of the solution $\mathbf{a}_{\epsilon_{\text{tol}}}$ obtained at the final iteration of Algorithm EC.1, for a tolerance parameter $\epsilon_{\text{tol}} > 0$. Let $P(0)$ be the optimal objective value of (P-constraint). Define, for any $\epsilon > 0$,

$$P(\epsilon) \triangleq \min_{\mathbf{a} \in \mathcal{A}} \left\{ g(\mathbf{a}) \;\middle|\; \sup_{(\mathbf{q}, \bar{\mathbf{q}}) \in \mathcal{U}_{\phi,h}(\mathbf{p})} -\sum_{k=1}^{N} \bar{q}_k u(x_k(\mathbf{a})) \leq c + \epsilon \right\}.$$

By construction, we have $P(\epsilon_{\text{tol}}) \leq g(\mathbf{a}_{\epsilon_{\text{tol}}})$ since $\mathbf{a}_{\epsilon_{\text{tol}}}$ is feasible for the problem of $P(\epsilon_{\text{tol}})$. Furthermore, the cutting-plane algorithm yields a lower bound on $P(0)$. Hence, $g(\mathbf{a}_{\epsilon_{\text{tol}}}) \leq P(0)$ for all $\epsilon_{\text{tol}} > 0$. Therefore,

$$0 \leq P(0) - g(\mathbf{a}_{\epsilon_{\text{tol}}}) \leq P(0) - P(\epsilon_{\text{tol}}).$$

Thus, it remains to bound $P(0) - P(\epsilon_{\text{tol}})$ from above. This can be done by utilizing the strong duality theorem, which is guaranteed by Assumption 4. Therefore, we have, for any $\epsilon > 0$,

$$\begin{aligned}
P(\epsilon) &= \sup_{\lambda \geq 0} \inf_{\mathbf{a} \in \mathcal{A}} g(\mathbf{a}) + \lambda \left(\sup_{(\mathbf{q}, \bar{\mathbf{q}}) \in \mathcal{U}_{\phi,h}(\mathbf{p})} - \sum_{k=1}^{N} \bar{q}_k u(x_k(\mathbf{a})) - c - \epsilon \right) \\
&\geq -\lambda^* \epsilon + \inf_{\mathbf{a} \in \mathcal{A}} g(\mathbf{a}) - \lambda^* \left(\sup_{(\mathbf{q}, \bar{\mathbf{q}}) \in \mathcal{U}_{\phi,h}(\mathbf{p})} - \sum_{k=1}^{N} \bar{q}_k u(x_k(\mathbf{a})) - c \right) \\
&= -\lambda^* \epsilon + P(0),
\end{aligned}$$

where λ^* is the KKT-vector of the problem $P(0)$ corresponding to the supremum constraint, which is a strictly positive constant (by Assumption 4) independent of ϵ. Hence, we have

$$P(0) - P(\epsilon_{\text{tol}}) \leq \lambda^* \epsilon_{\text{tol}}.$$

Therefore, the convergence follows as $\epsilon_{\text{tol}} \to 0$. □

Proof of Lemma 4. We first show that, for any $\mathbf{a} \in \mathcal{A}$, we have that

$$\sup_{(\mathbf{q}, \bar{\mathbf{q}}) \in \mathcal{U}_{\phi,h}(\mathbf{p})} \sum_{i=1}^{m} -\bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) = \sup_{(\mathbf{q}, \bar{\mathbf{q}}) \in \mathcal{U}_{\phi,h}^{(i_1,\ldots,i_m)}(\mathbf{p})} - \sum_{i=1}^{m} \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)), \tag{EC.6}$$

for any ranking $(i_1, \ldots, i_m) \in \mathcal{I}(\mathbf{a})$ as in Definition 2. We fix $\mathbf{a}$, denote $Y \triangleq f(\mathbf{a}, \mathbf{X})$ and consider the ranking $u(y_{(1)}) \geq \ldots \geq u(y_{(m)})$. By the definition of $\rho_{u,h,\mathbf{q}}(Y)$ in (4), we have that in our discrete setting $\rho_{u,h,\mathbf{q}}(Y)$ is equal to the rank-dependent sum

$$\rho_{u,h,\mathbf{q}}(Y) = -\sum_{i=1}^{m} h\left(\sum_{j=i}^{m} q_{(j)}\right) (u(y_{(i)}) - u(y_{(i-1)})),$$

where $-u(y_{(0)}) \triangleq 0$. We now claim that we have an equality between the rank-dependent sum above and the following optimization problem:

$$\begin{aligned}
\max_{\bar{\mathbf{q}} \in \mathbb{R}^m} \quad & \sum_{i=1}^{m} -\left(\sum_{j=i}^{m} \bar{q}_{(j)}\right) (u(y_{(i)}) - u(y_{(i-1)})) \\
\text{subject to} \quad & \sum_{j=i}^{m} \bar{q}_{(j)} \leq h\left(\sum_{j=i}^{m} q_{(j)}\right), \quad \forall i \in [m] \\
& \sum_{i=1}^{m} \bar{q}_i = 1 \\
& \bar{q}_i \geq 0, \quad \forall i \in [m].
\end{aligned}$$

Indeed, we can define $\bar{q}^*_{(i)} \triangleq h\left(\sum_{j=i}^{m} q_{(j)}\right) - h\left(\sum_{j=i+1}^{m} q_{(j)}\right)$, where $\bar{q}^*_{(m)} \triangleq h(q_{(m)})$. Then, $\sum_{j=i}^{m} \bar{q}^*_{(i)} = h\left(\sum_{j=i}^{m} q_{(j)}\right)$, and $\bar{\mathbf{q}}^*$ is a probability vector since $h(1) = 1$, $h(0) = 0$, and h is non-decreasing. Hence, $\bar{\mathbf{q}}^*$ is feasible for the above optimization problem. Furthermore, since

$-(u(y_{(i)})-u(y_{(i-1)}))\geq 0$ for all $i\geq 2$, the maximum is attained at a vector $\bar{\mathbf{q}}$ such that the constraint $\sum_{j=i}^{m}\bar{q}_{(i)}\leq h\left(\sum_{j=i}^{m}q_{(j)}\right)$ is an equality for all i, which uniquely defines $\bar{\mathbf{q}}^*$. An expansion of the alternating sum $\sum_{i=1}^{m}-\left(\sum_{j=i}^{m}q_{(j)}\right)(u(y_{(i)})-u(y_{(i-1)}))$ shows that it is also equal to the sum $\sum_{i=1}^{m}-q_{(i)}u(y_{(i)})$. Therefore, (EC.6) holds for all $\mathbf{a}\in\mathcal{A}$ and any ranking $(i_1,\ldots,i_m)\in\mathcal{I}(\mathbf{a})$. Let V^* be the optimal objective value of (P-constraint). Then, for any $\mathbf{a}_0$ such that $\mathcal{I}(\mathbf{a}_0)\subset\mathcal{I}(\mathbf{a}^*)$, we have that $U^*(\mathbf{a}_0)\leq V^*$, since $\mathbf{a}^*$ is feasible for the problem (21) by (EC.6). On the other hand, we also have the upper bound relation $V^*\leq U^*(\mathbf{a}_0)$, since $\mathcal{U}_{\phi,h}(\mathbf{p})\subset\mathcal{U}_{\phi,h}^{(i_1,\ldots,i_m)}(\mathbf{p})$ for any index vector $(i_1,\ldots,i_m)$. Hence, $U^*(\mathbf{a}_0)=V^*$. □

Proof of Theorem 5. Let $(\mathbf{q}_n^*,\bar{\mathbf{q}}_n^*)\in\operatorname{argmax}_{(\mathbf{q},\bar{\mathbf{q}})\in\mathcal{U}_{\phi,h}(\mathbf{p})}-\sum_{i=1}^{m}\bar{q}_i u(f(\mathbf{a}_n,\mathbf{x}_i))$. Denote $g(\mathbf{a},\mathbf{q},\bar{\mathbf{q}})=-\sum_{i=1}^{m}\bar{q}_i u(f(\mathbf{a},\mathbf{x}_i))$. Since the set $\mathcal{A}\times\mathcal{U}_{\phi,h}(\mathbf{p})$ is compact,[2] we may assume that there exists a limit $(\mathbf{a}_n,\mathbf{q}_n^*,\bar{\mathbf{q}}_n^*)\to(\mathbf{a}_L,\mathbf{q}_L,\bar{\mathbf{q}}_L)\in\mathcal{A}\times\mathcal{U}_{\phi,h}(\mathbf{p})$, as $n\to\infty$. We will show that $\mathbf{a}_L$ must be an optimal solution for (P-constraint). Indeed, since $(\mathbf{q}_n^*,\bar{\mathbf{q}}_n^*)$ are maximizers, we have that

$$c+\epsilon_{\mathrm{tol},n}\geq-\sum_{i=1}^{m}\bar{q}_{n,i}^*u(f(\mathbf{a}_n,\mathbf{x}_i))\geq\max_{(\mathbf{q},\bar{\mathbf{q}})\in\mathcal{U}_{\phi,h}(\mathbf{p})}-\sum_{i=1}^{m}\bar{q}_i u(f(\mathbf{a}_n,\mathbf{x}_i)).$$

Taking the limit as $n\to\infty$ yields

$$c\geq-\sum_{i=1}^{m}\bar{q}_{L,i}^*u(f(\mathbf{a}_L,\mathbf{x}_i))\geq\max_{(\mathbf{q},\bar{\mathbf{q}})\in\mathcal{U}_{\phi,h}(\mathbf{p})}-\sum_{i=1}^{m}\bar{q}_i u(f(\mathbf{a}_L,\mathbf{x}_i)).$$

Hence, by Theorem 1, this implies that $\mathbf{a}_L$ is feasible for (P-constraint). Since $\mathbf{a}_n$ is a sequence of solutions such that $g(\mathbf{a}_n)$ converges to the optimal objective value of (P-constraint) by Theorem 4, it follows from the continuity of g that $g(\mathbf{a}_L)$ equals to the optimal objective value of (P-constraint). Hence, $\mathbf{a}_L$ is an optimal solution of (P-constraint). Continuity of the functions $-u(f(\mathbf{a},\mathbf{x}_i))$ in $\mathbf{a}$, for all $i\in[m]$, implies that there exists some $N>0$, such that for all $n\geq N$, we have the inclusion of the ranking set: $\mathcal{I}(\mathbf{a}_n)\subset\mathcal{I}(\mathbf{a}_L)$. Therefore, Lemma 4 implies that $U^*(\mathbf{a}_n)$ converges to the optimal objective value of (P-constraint). □

Proof of Lemma 5. If $(\mathbf{q},\bar{\mathbf{q}},\mathbf{t})$ satisfies the constraints in (24), then $\bar{q}_i\leq l_j\cdot q_i+t_{i,j}$ for all $i\in[m]$ and $j\in[K]$. Then, by the non-negativity of the variables $t_{i,j}$, we also have that, for any subset $I\subset[m]$ and any $j\in[K]$,

$$\sum_{i\in I}\bar{q}_i\leq l_j\sum_{i\in I}q_i+\sum_{i\in I}t_{i,j}\leq l_j\sum_{i\in I}q_i+\sum_{i=1}^{m}t_{i,j}\leq l_j\sum_{i\in I}q_i+b_j.$$

Hence, $\sum_{i\in I}\bar{q}_i\leq\min_{j\in[K]}h_j(\sum_{i\in I}q_i)$ and thus $(\mathbf{q},\bar{\mathbf{q}})\in\mathcal{U}_{\phi,h}(\mathbf{p})$.

Conversely, let $(\mathbf{q},\bar{\mathbf{q}})\in\mathcal{U}_{\phi,h}(\mathbf{p})$. Then, $\bar{q}_i\leq l_j\cdot q_i+t_{i,j}$ for all $i\in[m]$ and $j\in[K]$ with $t_{i,j}\triangleq\max\{\bar{q}_i-l_jq_i,0\}$. Moreover, we have that, for all $j\in[K]$,

$$\sum_{i=1}^{m}t_{i,j}=\sum_{i=1}^{m}\max\{\bar{q}_i-l_jq_i,0\}=\sum_{i\in I_+}\bar{q}_i-\sum_{i\in I_+}l_jq_i\leq b_j,\ \text{where } I_+\triangleq\{i:\bar{q}_i-l_jq_i\geq 0\}.$$

Hence, $(\mathbf{q},\bar{\mathbf{q}},\mathbf{t})$ satisfies the constraints in (24). □

[2]The set $\mathcal{U}_{\phi,h}(\mathbf{p})$ is compact due to the lower-semicontinuity assumption made in Assumption 3.

Proof of Theorem 6. We reformulate the constraint

$$\max_{(\mathbf{q},\bar{\mathbf{q}},\mathbf{t})\in\bar{\mathcal{U}}} -\sum_{i=1}^{m} \bar{q}_i u(f(\mathbf{a},\mathbf{x}_i)) \leq c,$$

where

$$\bar{\mathcal{U}} \triangleq \left\{ (\mathbf{q},\bar{\mathbf{q}},\mathbf{t}) \in \mathbb{R}^{2m}_{\geq 0} \times \mathbb{R}^{mK}_{\geq 0} \;\middle|\; \begin{array}{l} \sum_{i=1}^m q_i = 1, \sum_{i=1}^m \bar{q}_i = 1,\ \sum_{i=1}^m p_i \phi\left(\frac{q_i}{p_i}\right) \leq r \\ \sum_{i=1}^m t_{i,j} \leq b_j,\ \forall j \in [K] \\ \bar{q}_i - l_j q_i \leq t_{i,j},\ \forall i \in [m],\ \forall j \in [K] \end{array} \right\}.$$

We note that $(\mathbf{p},\mathbf{p},\{\max\{(l_j-1)p_i,0\}_{i,j}\})$ is a Slater point in $\bar{\mathcal{U}}$ and that the left-hand side of the constraint above constitutes a maximization problem upper bounded by $c \in \mathbb{R}$. Therefore, strong duality holds, and we examine the Lagrangian function

$$\begin{aligned}
&L(\mathbf{a},\mathbf{q},\bar{\mathbf{q}},\alpha,\beta,\gamma,\lambda_{ij},t_{i,j},\nu_j) \\
&= -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a},\mathbf{x}_i)) - \alpha\left(\sum_{i=1}^m q_i - 1\right) - \beta\left(\sum_{i=1}^m \bar{q}_i - 1\right) - \gamma\left(\sum_{i=1}^m p_i\phi\left(\frac{q_i}{p_i}\right) - r\right) \\
&\quad - \sum_{i=1}^m\sum_{j=1}^K \lambda_{ij}(\bar{q}_i - l_j q_i - t_{i,j}) - \sum_{j=1}^K \nu_j\left(\sum_{i=1}^m t_{i,j} - b_j\right) \\
&= \alpha + \beta + \gamma r + \sum_{j=1}^K \nu_j b_j - \sum_{i=1}^m \left(u(f(\mathbf{a},\mathbf{x}_i)) + \beta + \sum_{j=1}^K \lambda_{ij}\right)\bar{q}_i \\
&\quad + \sum_{i=1}^m p_i\left(\left(-\alpha + \sum_{j=1}^K \lambda_{ij} l_j\right)\frac{q_i}{p_i} - \gamma\phi\left(\frac{q_i}{p_i}\right)\right) + \sum_{i=1}^m\sum_{j=1}^K (\lambda_{ij} - \nu_j)t_{i,j}.
\end{aligned}$$

We have

$$\begin{aligned}
&\sup_{\bar{q}_1,\ldots,\bar{q}_m\geq 0} -\sum_{i=1}^m \left(u(f(\mathbf{a},\mathbf{x}_i)) + \beta + \sum_{j=1}^K \lambda_{ij}\right)\bar{q}_i \\
&= \begin{cases} 0 & \text{if } u(f(\mathbf{a},\mathbf{x}_i)) + \beta + \sum_{j=1}^K \lambda_{ij} \geq 0,\ \forall i \in [m] \\ \infty & \text{else,} \end{cases}
\end{aligned} \tag{EC.7}$$

$$\sup_{q_1,\ldots,q_m\geq 0} \sum_{i=1}^m p_i\left(\left(-\alpha + \sum_{j=1}^K \lambda_{ij} l_j\right)\frac{q_i}{p_i} - \gamma\phi\left(\frac{q_i}{p_i}\right)\right) = \sum_{i=1}^m p_i\gamma\phi^*\left(\frac{-\alpha + \sum_{j=1}^K \lambda_{ij} l_j}{\gamma}\right),$$

and

$$\sup_{t_{1,1},\ldots,t_{m,K}\geq 0} \sum_{i=1}^m\sum_{j=1}^K (\lambda_{ij} - \nu_j)t_{i,j} = \begin{cases} 0 & \text{if } \lambda_{ij} \leq \nu_j,\ \forall i \in [m],\ \forall j \in [K] \\ \infty & \text{else.} \end{cases} \tag{EC.8}$$

Therefore, employing the same arguments as in the proof of Theorem 2, the reformulated robust counterpart is given by

$$\begin{cases} \alpha + \beta + \gamma r + \sum_{j=1}^{K} \nu_j b_j + \sum_{i=1}^{m} p_i \gamma \phi^* \left(\frac{-\alpha + \sum_{j=1}^{K} \lambda_{ij} l_j}{\gamma} \right) \leq c \\ -u(f(\mathbf{a}, \mathbf{x}_i)) - \beta - \sum_{j=1}^{K} \lambda_{ij} \leq 0, \ \forall i \in [m] \\ \lambda_{ij} \leq \nu_j, \ \forall i \in [m], \ \forall j \in [K] \\ \alpha, \beta \in \mathbb{R}, \gamma, \lambda_{ij}, \nu_j \geq 0. \end{cases}$$

In the nominal case where $\mathbf{q} = \mathbf{p}$, we have to reformulate the following constraint:

$$\max_{(\bar{\mathbf{q}}, \mathbf{t}) \in \bar{\mathcal{U}}_{\text{nom}}} - \sum_{i=1}^{m} \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) \leq c,$$

where

$$\bar{\mathcal{U}}_{\text{nom}} \triangleq \left\{ (\bar{\mathbf{q}}, \mathbf{t}) \in \mathbb{R}_{\geq 0}^{m} \times \mathbb{R}_{\geq 0}^{mK} \;\middle|\; \begin{array}{l} \sum_{i=1}^{m} \bar{q}_i = 1 \\ \sum_{i=1}^{m} t_{i,j} \leq b_j, \ \forall j \in [K] \\ \bar{q}_i - l_j p_i \leq t_{i,j}, \ \forall i \in [m], \ \forall j \in [K] \end{array} \right\}.$$

The above maximization problem is a linear programming problem bounded from above. Therefore, strong duality applies, and we examine the Lagrangian function

$$\begin{aligned} & L(\mathbf{a}, \bar{\mathbf{q}}, \beta, \lambda_{ij}, t_{i,j}, \nu_j) \\ &= - \sum_{i=1}^{m} \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) - \beta \left(\sum_{i=1}^{m} \bar{q}_i - 1 \right) - \sum_{i=1}^{m} \sum_{j=1}^{K} \lambda_{ij} (\bar{q}_i - l_j p_i - t_{i,j}) - \sum_{j=1}^{K} \nu_j \left(\sum_{i=1}^{m} t_{i,j} - b_j \right) \\ &= \beta + \sum_{j=1}^{K} \nu_j b_j + \sum_{i=1}^{m} \sum_{j=1}^{K} \lambda_{ij} l_j p_i + \sum_{i=1}^{m} \bar{q}_i \left(-u(f(\mathbf{a}, \mathbf{x}_i)) - \beta - \sum_{j=1}^{K} \lambda_{ij} \right) + \sum_{i=1}^{m} \sum_{j=1}^{K} (\lambda_{ij} - \nu_j) t_{i,j}. \end{aligned}$$

By (EC.7)–(EC.8), we have that

$$\begin{aligned} & \inf_{\substack{\lambda_{ij}, \nu_j \geq 0 \\ \beta \in \mathbb{R}}} \sup_{\bar{\mathbf{q}}, \mathbf{t} \geq \mathbf{0}} L(\mathbf{a}, \bar{\mathbf{q}}, \beta, \lambda_{ij}, t_{i,j}, \nu_j) \\ &= \inf_{\substack{\lambda_{ij}, \nu_j \geq 0 \\ \beta \in \mathbb{R}}} \left\{ \beta + \sum_{j=1}^{K} \nu_j b_j + \sum_{i=1}^{m} \sum_{j=1}^{K} \lambda_{ij} l_j p_i \;\middle|\; -u(f(\mathbf{a}, \mathbf{x}_i)) - \beta - \sum_{j=1}^{K} \lambda_{ij} \leq 0, \forall i, \ \lambda_{ij} \leq \nu_j, \forall i, j \right\}. \end{aligned}$$

Hence, the reformulated robust counterpart is given by

$$\begin{cases} \beta + \sum_{j=1}^{K} \nu_j b_j + \sum_{i=1}^{m} \sum_{j=1}^{K} \lambda_{ij} l_j p_i \leq c \\ -u(f(\mathbf{a}, \mathbf{x}_i)) - \beta - \sum_{j=1}^{K} \lambda_{ij} \leq 0, \ \forall i \in [m] \\ \lambda_{ij} \leq \nu_j, \ \forall i \in [m], \ \forall j \in [K] \\ \lambda_{ij}, \nu_j \geq 0, \ \forall i \in [m], \ \forall j \in [K]. \end{cases}$$

□

Proof of Lemma 6. First, we have that

$$e_i(x_{i+1}) = \sup_{x\in[x_i,1]} \left\{ h(x) - \frac{h(x_{i+1}) - h(x_i)}{x_{i+1} - x_i}(x - x_i) - h(x_i) \right\},$$

since, by concavity, the supremum is only taken in the interval $[x_i, x_{i+1}]$. Therefore, we can extend the feasibility region to $[x_i, 1]$.

Next, again by concavity, we have that $-\frac{h(x_{i+1})-h(x_i)}{x_{i+1}-x_i}(x - x_i)$ is an increasing function of x_{i+1}, for all $x \in (x_i, 1]$. Therefore, $e_i(x_{i+1})$ is increasing in x_{i+1}. It is also a continuous function in x_{i+1}. This is because the function

$$\tilde{e}_i(y) \triangleq \sup_{x\in[x_i,1]} \{h(x) - y(x - x_i) - h(x_i)\},$$

is convex in y; indeed, it is a supremum of a linear function of y. Hence, $\tilde{e}_i$ is continuous on the interior of its domain, which is the whole of $\mathbb{R}$, because $\tilde{e}_i(y)$ is a supremum of a continuous function on a compact interval. Thus, $\tilde{e}_i(y)$ exists and is finite everywhere. This implies that e_i is continuous for all $x_{i+1} \in (x_i, 1]$, since $e_i(x_{i+1}) = \tilde{e}_i\left(\frac{h(x_{i+1})-h(x_i)}{x_{i+1}-x_i}\right)$ is a composition of continuous functions.

Finally, we show the existence of a x_{i+1} such that $e_i(x_{i+1}) = \epsilon$, for any given $\epsilon > 0$, under the assumption that $e_i(1) > \epsilon$. This is guaranteed if we can find a $z \in (x_i, 1)$ such that $e_i(z) < \epsilon$. Since h has decreasing slope and is increasing, we have that for any $y \in (x_i, 1)$ and all $x \in [x_i, y]$:

$$h(x) - \frac{h(y) - h(x_i)}{y - x_i}(x - x_i) - h(x_i) \le h(y) - h(x_i).$$

We note that the maximization problem in $e_i(y)$ can be restricted to the interval $[x_i, y]$. Therefore, $e_i(y) \le h(y) - h(x_i)$. Taking $y \downarrow x_i$, it follows by the continuity of h that such a point z must exist. □

Proof of Theorem 7 (Part I). We split the proof into two parts. We first treat the case of problem (P) and next the case of problem (P-constraint). Fix an $\mathbf{a} \in \mathcal{A}$ and consider the ranked realizations $u(f(\mathbf{a}, \mathbf{x}_{(1)})) \ge \ldots \ge u(f(\mathbf{a}, \mathbf{x}_{(m)}))$. Let h_1, h_2 be any two concave distortion functions such that $\sup_{t\in[0,1]} |h_1(t) - h_2(t)| \le \epsilon$. Set $u(f(\mathbf{a}, \mathbf{x}_{(0)})) \triangleq 0$. Then, we have

$$\begin{aligned}
\rho_{u,h_1,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) &= \sum_{i=1}^{m} h_1\left(\sum_{j=i}^{m} q_{(j)}\right)(u(f(\mathbf{a}, \mathbf{x}_{(i-1)})) - u(f(\mathbf{a}, \mathbf{x}_{(i)}))) \\
&\le -u(f(\mathbf{a}, \mathbf{x}_{(1)})) + \sum_{i=2}^{m}\left(h_2\left(\sum_{j=i}^{m} q_{(j)}\right) + \epsilon\right)(u(f(\mathbf{a}, \mathbf{x}_{(i-1)})) - u(f(\mathbf{a}, \mathbf{x}_{(i)}))) \\
&= \rho_{u,h_2,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) + \epsilon \cdot \left(\max_{i\in[m]} u(f(\mathbf{a}, \mathbf{x}_i)) - \min_{i\in[m]} u(f(\mathbf{a}, \mathbf{x}_i))\right) \\
&\le \rho_{u,h_2,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) + \epsilon \cdot \left(M + \max_{i\in[m]} -u(f(\mathbf{a}, \mathbf{x}_i))\right),
\end{aligned}$$

where $M \triangleq \sup_{\mathbf{a}\in\mathcal{A}, i\in[m]} u(f(\mathbf{a}, \mathbf{x}_i)) < \infty$. The idea now is to bound the term $\max_{i\in[m]} -u(f(\mathbf{a}, \mathbf{x}_i))$ on a subset $\mathcal{A}_0 \subset \mathcal{A}$ of the feasible region, for which the restriction of the minimization problem (P) on $\mathcal{A}_0$ does not change the original optimal value.

We take $h_1 \equiv h$ and $h_2 \in \{h_\epsilon, \tilde{h}_\epsilon\}$. By Assumption 1, there exists $\mathbf{a}_0 \in \mathcal{A}$ such that $\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a}_0, \mathbf{X})) < \infty$. This implies $\rho_{u,h,\mathbf{p}}(f(\mathbf{a}_0, \mathbf{X})) < \infty$, hence $\max_{i\in[m]} -u(f(\mathbf{a}_0, \mathbf{x}_i)) < \infty$. Then, we also have that

$$\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_2,\mathbf{q}}(f(\mathbf{a}_0, \mathbf{X})) \leq \max_{i\in[m]} -u(f(\mathbf{a}_0, \mathbf{x}_i)) < \infty.$$

Therefore, we may define the finite number

$$c_0 \triangleq \max_{j=1,2} \left\{ \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_j,\mathbf{q}}(f(\mathbf{a}_0, \mathbf{X})) \right\} < \infty.$$

Then, for all $\mathbf{a} \in \mathcal{A}$ such that $\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_j,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) \leq c_0$ for any $j \in \{1, 2\}$, we have that

$$-\sum_{i=1}^{m} p_i u(f(\mathbf{a}, \mathbf{x}_i)) \leq \rho_{u,h_j,\mathbf{p}}(f(\mathbf{a}, \mathbf{X})) \leq c_0, \tag{EC.9}$$

due to concavity (which implies $h_j(x) \geq x, \forall x \in [0, 1]$, for any $j \in \{1, 2\}$). Therefore, for all such $\mathbf{a}$, we have,

$$\sup_{i\in[m]} -u(f(\mathbf{a}, \mathbf{x}_i)) \leq \frac{c_0 + M}{p_{\min}}, \tag{EC.10}$$

as shown in the proof of Theorem 3. Define,

$$\mathcal{A}_0 \triangleq \left\{ \mathbf{a} \in \mathcal{A} \;\middle|\; \sup_{i\in[m]} -u(f(\mathbf{a}, \mathbf{x}_i)) \leq \frac{c_0 + M}{p_{\min}} \right\}.$$

Then, we have that for any $j = 1, 2$:

$$\min_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_j,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) = \min_{\mathbf{a}\in\mathcal{A}_0} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_j,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})),$$

since any potential minimizer $\mathbf{a}$ satisfies $\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_j,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) \leq c_0$ and thus (EC.10). Therefore,

$$\begin{aligned}
\min_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_1,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) &= \min_{\mathbf{a}\in\mathcal{A}_0} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_1,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) \\
&\leq \min_{\mathbf{a}\in\mathcal{A}_0} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_2,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) + \epsilon \cdot \left(M + \frac{c_0 + M}{p_{\min}} \right) \\
&= \min_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_2,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) + \epsilon \cdot \left(M + \frac{c_0 + M}{p_{\min}} \right).
\end{aligned}$$

By symmetry, we thus have

$$\left| \min_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_1,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) - \min_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_2,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) \right| \leq \epsilon \cdot \left(M + \frac{c_0 + M}{p_{\min}} \right),$$

which approaches zero as $\epsilon \to 0$. $\square$

Proof of Theorem 7 (Part II). We define

$$L_\epsilon \triangleq \min_{\mathbf{a}\in\mathcal{A}} \left\{ g(\mathbf{a}) \;\middle|\; \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_\epsilon,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \le c \right\},$$
$$U_\epsilon \triangleq \min_{\mathbf{a}\in\mathcal{A}} \left\{ g(\mathbf{a}) \;\middle|\; \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,\tilde{h}_\epsilon,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \le c \right\},$$

and let P_0 denote the optimal objective value of (P-constraint). By definition, $L_\epsilon \le P_0 \le U_\epsilon$.

We first note that L_ϵ has a nonempty feasible set, which follows from Assumption 4 and the fact that $\rho_{u,h_\epsilon,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \le \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X}))$ for all $\mathbf{a}\in\mathcal{A}$ and any probability vector $\mathbf{q}$. Then, following the first part of the proof of Theorem 7, we can show that for any $\mathbf{a}\in\mathcal{A}$ that is feasible for L_ϵ, we have that, for all $\mathbf{q}$,

$$\rho_{u,h_\epsilon,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \ge \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) - \epsilon\left(M + \frac{c+M}{p_{\min}}\right).$$

Therefore, we also have the following implication for any $\mathbf{a}\in\mathcal{A}$:

$$\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_\epsilon,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \le c \Rightarrow \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \le c + \epsilon\left(M + \frac{c+M}{p_{\min}}\right).$$

Hence,

$$\begin{aligned}
L_\epsilon &\ge \min_{\mathbf{a}\in\mathcal{A}} \left\{ g(\mathbf{a}) \;\middle|\; \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \le c + \epsilon\left(M + \frac{c+M}{p_{\min}}\right) \right\} \\
&\overset{(*)}{\ge} \sup_{\lambda\ge 0} \inf_{\mathbf{a}\in\mathcal{A}} g(\mathbf{a}) + \lambda\left(\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) - c - \epsilon\left(M + \frac{c+M}{p_{\min}}\right)\right) \\
&\ge \inf_{\mathbf{a}\in\mathcal{A}} g(\mathbf{a}) + \lambda^*\left(\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) - c - \epsilon\left(M + \frac{c+M}{p_{\min}}\right)\right) \\
&\overset{(**)}{=} P_0 - \lambda^*\cdot\epsilon\left(M + \frac{c+M}{p_{\min}}\right),
\end{aligned}$$

where in $(*)$ we used weak duality and in $(**)$ we used λ^*, the KKT-vector of (P-constraint) corresponding to the constraint $\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \le c$, which is a strictly positive constant by Assumption 4 and does not depend on ϵ. Therefore, we have

$$0 \le P_0 - L_\epsilon \le \lambda^*\epsilon\left(M + \frac{c+M}{p_{\min}}\right),$$

which approaches zero as $\epsilon \to 0$.

Similarly, we have the following implication for any $\mathbf{a}\in\mathcal{A}$:

$$\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \le c - \epsilon\left(M + \frac{c+M}{p_{\min}}\right) \Rightarrow \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,\tilde{h}_\epsilon,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \le c.$$

Therefore,

$$U_\epsilon \le \min_{\mathbf{a}\in\mathcal{A}} \left\{ g(\mathbf{a}) \;\middle|\; \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \le c - \epsilon\left(M + \frac{c+M}{p_{\min}}\right) \right\}. \tag{EC.11}$$

We note that the minimization problem on the right-hand side of (EC.11) contains a Slater point, for ϵ sufficiently small. Indeed, by Assumption 4, there exists a point $\mathbf{a}_0 \in \mathrm{int}(\mathcal{A})$, such that

$$\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a}_0,\mathbf{X})) \triangleq c_0 < c.$$

Then, for $\epsilon \leq \frac{c-c_0}{2\left(M+\frac{c+M}{p_{\min}}\right)}$, we have that

$$\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a}_0,\mathbf{X})) < c - \epsilon\left(M + \frac{c+M}{p_{\min}}\right). \tag{EC.12}$$

Therefore, together with Assumption 4, as well as the reformulation in Theorem 1, we may apply the strong duality theorem and obtain the upper estimation

$$\begin{aligned}
U_\epsilon &\leq \sup_{\lambda\geq 0}\min_{\mathbf{a}\in\mathcal{A}} g(\mathbf{a}) + \lambda\left(\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) - c + \epsilon\left(M + \frac{c+M}{p_{\min}}\right)\right)\\
&= \min_{\mathbf{a}\in\mathcal{A}} g(\mathbf{a}) + \lambda^*(\epsilon)\left(\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) - c + \epsilon\left(M + \frac{c+M}{p_{\min}}\right)\right)\\
&\leq \lambda^*(\epsilon)\cdot\epsilon\left(M + \frac{c+M}{p_{\min}}\right) + \sup_{\lambda\geq 0}\min_{\mathbf{a}\in\mathcal{A}} g(\mathbf{a}) + \lambda\left(\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) - c\right)\\
&= \lambda^*(\epsilon)\cdot\epsilon\left(M + \frac{c+M}{p_{\min}}\right) + P_0,
\end{aligned}$$

where $\lambda^*(\epsilon)$ is the KKT-vector of the minimization problem on the right-hand side of (EC.11), corresponding to the supremum constraint, which depends on ϵ. As a final step, we will show that $\lambda^*(\epsilon)$ can be further bounded by a constant that does not depend on ϵ, for ϵ sufficiently small. Indeed, let $\mathbf{a}_0$ be the point in (EC.12) and let $\epsilon \leq \frac{c-c_0}{2\left(M+\frac{c+M}{p_{\min}}\right)}$. Since $P_0 \leq U_\epsilon$,

$$\begin{aligned}
P_0 &\leq \min_{\mathbf{a}\in\mathcal{A}} g(\mathbf{a}) + \lambda^*(\epsilon)\left(\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) - c + \epsilon\left(M + \frac{c+M}{p_{\min}}\right)\right)\\
&\leq g(\mathbf{a}_0) + \lambda^*(\epsilon)\left(\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a}_0,\mathbf{X})) - c + \epsilon\left(M + \frac{c+M}{p_{\min}}\right)\right)\\
&\leq g(\mathbf{a}_0) - \lambda^*(\epsilon)\frac{c-c_0}{2},
\end{aligned}$$

which implies that, for all $\epsilon \leq \frac{c-c_0}{2\left(M+\frac{c+M}{p_{\min}}\right)}$, we have

$$\lambda^*(\epsilon) \leq \frac{2(g(\mathbf{a}_0) - P_0)}{c - c_0} \triangleq C^*.$$

Hence, for ϵ sufficiently small,

$$0 \leq U_\epsilon - P_0 \leq C^*\cdot\epsilon\left(M + \frac{c+M}{p_{\min}}\right),$$

which converges to zero as $\epsilon \to 0$. □

Proof of Proposition 1. Choose the function $h(p) = p^2$. Then $h(0) = 0$, $h(1) = 1$ and $h(p) < p,\ \forall p \in (0,1)$. Due to the compactness of $\mathcal{U}$, there exists a maximizer $\mathbf{q}^*$ of $\sup_{\mathbf{q}\in\mathcal{U}} \rho_{h,\mathbf{q}}(X)$. Let $\hat{h}(p) = p$ be the concave envelope of h. By the assumptions on $\mathcal{U}$, we have that $\mathbf{q}_1^* < 1$. Hence,

$$\rho_{\hat{h},\mathbf{q}^*}(X) - \rho_{h,\mathbf{q}^*}(X) = \sum_{i=2}^{m} \left(\hat{h}\left(\sum_{k=i}^{m} q_k^*\right) - h\left(\sum_{k=i}^{m} q_k^*\right)\right)(x_{i-1} - x_i) > 0,$$

since $\hat{h}(p) > h(p)$ for all $p \in (0,1)$ and $x_{i-1} - x_i > 0$ for all $i \in [m]$. Therefore, $\sup_{\mathbf{q}\in\mathcal{U}} \rho_{\hat{h},\mathbf{q}}(X) > \sup_{\mathbf{q}\in\mathcal{U}} \rho_{h,\mathbf{q}}(X)$. □

Proof of Theorem 8. It is sufficient to show that for any X, with outcomes $\{x_i\}_{i=1}^m$, we have

$$\rho_{h,\mathbf{p}}(X) = \sup_{\mathbf{q}\in M_h^{\mathrm{ca}}(\mathbf{p})} \sum_{i=1}^{m} -q_i x_i - \sup_{\bar{\mathbf{q}}\in N_{\bar{h}}^{\mathrm{cv}}(\mathbf{p})} \sum_{i=1}^{m} \bar{q}_i x_i,$$

since $\rho_{u,h,\mathbf{p}}(f(\mathbf{a},\mathbf{X})) = \rho_{h,\mathbf{p}}(u(f(\mathbf{a},\mathbf{X}))$. Let $-x_{(1)} \geq \ldots \geq -x_{(m)}$ be the ranked realizations of X. Let k^* be the index where $h\left(\sum_{s=1}^{k^*} p_{(s)}\right) \leq h(p^0) \leq h\left(\sum_{s=1}^{k^*+1} p_{(s)}\right)$. Then,

$$\begin{aligned}
&\rho_{h,\mathbf{p}}(X)\\
&= \sum_{i=1}^{k^*}\left(h\left(\sum_{s=1}^{i} p_{(s)}\right) - h\left(\sum_{s=1}^{i-1} p_{(s)}\right)\right)\cdot(-x_{(i)}) + \left(h(p^0) - h\left(\sum_{s=1}^{k^*} p_{(s)}\right)\right)\cdot(-x_{(k^*+1)})\\
&\quad + \left(h\left(\sum_{s=1}^{k^*+1} p_{(s)}\right) - h(p^0)\right)\cdot(-x_{(k^*+1)}) + \sum_{i=k^*+2}^{m}\left(h\left(\sum_{s=1}^{i} p_{(s)}\right) - h\left(\sum_{s=1}^{i-1} p_{(s)}\right)\right)\cdot(-x_{(i)}),
\end{aligned}$$

where an empty sum $\sum_{s=1}^{0}$ is zero by convention. We can also express the second line of the previous expression in terms of the dual function $\bar{h}(p) \triangleq 1 - h(1-p)$, and rearrange the indices to obtain:

$$\begin{aligned}
&\sum_{i=k^*+2}^{m}\left(h\left(\sum_{s=1}^{i} p_{(s)}\right) - h\left(\sum_{s=1}^{i-1} p_{(s)}\right)\right)\cdot(-x_{(i)}) + \left(h\left(\sum_{s=1}^{k^*+1} p_{(s)}\right) - h(p^0)\right)\cdot(-x_{(k^*+1)})\\
&= \sum_{i=k^*+2}^{m}\left(\bar{h}\left(\sum_{s=i}^{m} p_{(s)}\right) - \bar{h}\left(\sum_{s=i+1}^{m} p_{(s)}\right)\right)\cdot(-x_{(i)}) + \left(\bar{h}(1-p^0) - \bar{h}\left(\sum_{s=k^*+2}^{m} p_{(s)}\right)\right)\cdot(-x_{(k^*+1)})\\
&= -\sum_{i=1}^{m-k^*-1}\left(\bar{h}\left(\sum_{s=1}^{i} p_{(m-s+1)}\right) - \bar{h}\left(\sum_{s=1}^{i-1} p_{(m-s+1)}\right)\right)\cdot x_{(m-i+1)}\\
&\qquad - \left(\bar{h}(1-p^0) - \bar{h}\left(\sum_{s=1}^{m-k^*-1} p_{(m-s+1)}\right)\right)\cdot x_{(k^*+1)},
\end{aligned}$$

where the empty sum $\sum_{s=m+1}^{m}$ is again zero. It remains to show that the sums are equal to their dual representations:

$$\begin{aligned}
\sup_{\mathbf{q}\in M_h^{\mathrm{ca}}(\mathbf{p})} \sum_{i=1}^{m} -q_i x_i = \sum_{i=1}^{k^*}&\left(h\left(\sum_{s=1}^{i} p_{(s)}\right) - h\left(\sum_{s=1}^{i-1} p_{(s)}\right)\right)\cdot(-x_{(i)})\\
&+ \left(h(p^0) - h\left(\sum_{s=1}^{k^*} p_{(s)}\right)\right)\cdot(-x_{(k^*+1)}),
\end{aligned} \tag{EC.13}$$

and

$$\sup_{\bar{\mathbf{q}}\in N_{\bar{h}}^{\mathrm{cv}}(\mathbf{p})}\sum_{i=1}^{m}\bar{q}_i x_i = \sum_{i=1}^{m-k^*-1}\left(\bar{h}\left(\sum_{s=1}^{i}p_{(m-s+1)}\right)-\bar{h}\left(\sum_{s=1}^{i-1}p_{(m-s+1)}\right)\right)\cdot x_{(m-i+1)} + \left(\bar{h}(1-p^0)-\bar{h}\left(\sum_{s=1}^{m-k^*-1}p_{(m-s+1)}\right)\right)\cdot x_{(k^*+1)}. \tag{EC.14}$$

We will first show (EC.13); (EC.14) follows similarly. Let h_0 be defined as follows:

$$h_0(p)\triangleq\begin{cases}h(p) & p\le p^0\\ h(p^0) & p\ge p^0.\end{cases} \tag{EC.15}$$

Then, h_0 is non-decreasing and concave. Let P be the probability measure induced by the vector $(p_i)_{i=1}^m$. Then, by Lemma EC.1, $\mu_1\triangleq h_0\circ P$ is a monotone, submodular set function. By the definition of a Choquet integral, we have

$$\begin{aligned}&\int -X\mathrm{d}\mu_1\\ &=\sum_{i=1}^{k^*}\left(h_0\left(\sum_{s=1}^{i}p_{(s)}\right)-h_0\left(\sum_{s=1}^{i-1}p_{(s)}\right)\right)\cdot(-x_{(i)})+\left(h_0(p^0)-h_0\left(\sum_{s=1}^{k^*}p_{(s)}\right)\right)\cdot(-x_{(k^*+1)})\\ &=\sum_{i=1}^{k^*}\left(h\left(\sum_{s=1}^{i}p_{(s)}\right)-h\left(\sum_{s=1}^{i-1}p_{(s)}\right)\right)\cdot(-x_{(i)})+\left(h(p^0)-h\left(\sum_{s=1}^{k^*}p_{(s)}\right)\right)\cdot(-x_{(k^*+1)}).\end{aligned}$$

Moreover, since μ_1 is monotone and submodular, we have by Proposition 10.3 of Denneberg (1994),

$$\int -X\mathrm{d}\mu_1=\left\{\sum_{i=1}^{m}-q_i x_i \;\middle|\; q_i\ge 0,\sum_{i\in J}q_i\le h_0\left(\sum_{i\in J}p_i\right),\ \forall J\subset[m],\ \sum_{i=1}^{m}q_i=h_0(p^0)\right\},$$

which involves the same feasible set as in (29).

Similarly, define the function

$$\bar{h}_0(p)\triangleq\begin{cases}\bar{h}(p) & p\le 1-p^0\\ \bar{h}(1-p^0) & p\ge 1-p^0.\end{cases} \tag{EC.16}$$

Then, $\bar{h}_0$ is concave and non-decreasing. Hence, by the same arguments as above, we have that $\mu_2\triangleq\bar{h}_0\circ P$ is monotone and submodular. Therefore,

$$\int X\mathrm{d}\mu_2=\sum_{i=1}^{m-k^*-1}\left(\bar{h}\left(\sum_{s=1}^{i}p_{(m-s+1)}\right)-\bar{h}\left(\sum_{s=1}^{i-1}p_{(m-s+1)}\right)\right)\cdot x_{(m-i+1)} + \left(\bar{h}(1-p^0)-\bar{h}\left(\sum_{s=1}^{m-k^*-1}p_{(s)}\right)\right)\cdot x_{(k^*+1)}.$$

Moreover, we have

$$\int X\mathrm{d}\mu_2=\sup\left\{\sum_{i=1}^{m}\bar{q}_i x_i \;\middle|\; \bar{q}_i\ge 0,\sum_{i\in J}\bar{q}_i\le \bar{h}_0\left(\sum_{i\in J}p_i\right),\ \forall J\subset[m],\ \sum_{i=1}^{m}\bar{q}_i=\bar{h}_0(1-p^0)\right\},$$

which implies (30). $\square$

Lemma EC.1. *Let Ω be a set and let $\mathcal{F} = 2^{\Omega}$ be the collection of all subsets of Ω. Furthermore, let $h : [0,1] \to \mathbb{R}$ be a non-decreasing concave function such that $h(0) = 0$ and let $P : \mathcal{F} \to [0, \infty)$ be an additive set function. Then, $\mu \triangleq h \circ P$ is a monotone, submodular set function.*

Proof. Monotonicity of μ is clear due to the additivity of P and the monotonicity of h. Let $A, B \in \mathcal{F}$. Additivity of P also implies that

$$P(A \cup B) + P(A \cap B) = P(A) + P(B).$$

Let $a \triangleq P(A),\ b \triangleq P(B)$. Then, we have

$$P(A \cap B) \triangleq i \leq a \leq b \leq u \triangleq P(A \cup B).$$

Since $b - i = u - a$, we have that by concavity of h,

$$\frac{h(b) - h(i)}{b - i} \geq \frac{h(u) - h(a)}{u - a},$$

which, via $h(i) + h(u) \leq h(a) + h(b)$, implies the submodularity of μ. $\square$

Proof of Theorem 9. By Theorem 8, it is sufficient to show that the sets $M_h^{\mathrm{ca}}(\mathbf{p})$ and $N_{\bar{h}}^{\mathrm{cv}}(\mathbf{p})$ are the projection of the following sets on the coordinates $\mathbf{q}$ and $\bar{\mathbf{q}}$, respectively:

$$M_{h,l}^{\mathrm{ca}}(\mathbf{p}) = \left\{ \mathbf{q} \in \mathbb{R}^m, \mathbf{t}^{(1)} \in \mathbb{R}^{m \times K_1} \;\middle|\; \begin{array}{l} q_i \geq 0 \\ q_i \leq l_k^{(1)} p_i + t_{ik}^{(1)},\ \forall i \in [m],\ \forall k \in [K_1] \\ \sum_{i=1}^m t_{ik}^{(1)} \leq b_k^{(1)},\ \forall k \in [K_1] \\ \sum_{i=1}^m q_i = h(p^0) \end{array} \right\}, \tag{EC.17}$$

and

$$N_{\bar{h},l}^{\mathrm{cv}}(\mathbf{p}) = \left\{ \bar{\mathbf{q}} \in \mathbb{R}^m, \mathbf{t}^{(2)} \in \mathbb{R}^{m \times K_2} \;\middle|\; \begin{array}{l} \bar{q}_i \geq 0 \\ \bar{q}_i \leq l_k^{(2)} p_i + t_{ik}^{(2)},\ \forall i \in [m],\ \forall k \in [K_2] \\ \sum_{i=1}^m t_{ik}^{(2)} \leq b_k^{(2)}, \forall k \in [K_2] \\ \sum_{i=1}^m \bar{q}_i = \bar{h}(1 - p^0) \end{array} \right\}, \tag{EC.18}$$

We will only show this for $M_{h,l}^{\mathrm{ca}}(\mathbf{p})$, since the case for $N_{\bar{h},l}^{\mathrm{cv}}(\mathbf{p})$ is identical, *mutatis mutandis.*

Let $(\mathbf{q}, \mathbf{t}^{(1)}) \in M_{h,l}^{\mathrm{ca}}(\mathbf{p})$. Then, $q_i \leq l_k^{(1)} \cdot p_i + t_{ik}^{(1)}$ for all $i \in [m]$ and $k \in [K_1]$. Thus, for any subset $I \subset [m]$, we also have that, for all $j \in [K_1]$,

$$\sum_{i \in I} q_i \leq l_k^{(1)} \sum_{i \in I} p_i + \sum_{i \in I} t_{ik}^{(1)} \leq l_k^{(1)} \sum_{i \in I} p_i + \sum_{i=1}^m t_{ik}^{(1)} \leq l_k^{(1)} \sum_{i \in I} p_i + b_k^{(1)}.$$

Hence, $\mathbf{q} \in M_h^{\mathrm{ca}}(\mathbf{p})$. Conversely, let $\mathbf{q} \in M_h^{\mathrm{ca}}(\mathbf{p})$. Then, $q_i \leq l_k^{(1)} \cdot p_i + t_{ik}^{(1)}$ for all $i \in [m]$ and $k \in [K_1]$ with $t_{ik}^{(1)} \triangleq \max\{q_i - l_k^{(1)} p_i, 0\}$. Moreover, we have, for all $k \in [K_1]$,

$$\sum_{i=1}^m t_{ik}^{(1)} = \sum_{i=1}^m \max\{q_i - l_k^{(1)} p_i, 0\} = \sum_{i \in I_+} q_i - \sum_{i \in I_+} l_k^{(1)} p_i \leq b_k^{(1)}, \text{ where } I_+ \triangleq \{i : q_i - l_k^{(1)} p_i \geq 0\}.$$

Indeed, the last inequality follows from the fact that, for any $\mathbf{q} \in M_h^{\mathrm{ca}}(\mathbf{p})$, we have that $\sum_{i \in I} \bar{q}_i \leq l_k^{(1)} \sum_{i \in I} q_i + b_k^{(1)}$ if $\sum_{i \in I} p_i \leq h(p^0)$. For any index set I such that $\sum_{i \in I} p_i > h(p^0)$, we also have that $l_k^{(1)} \sum_{i\in I} p_i + b_k^{(1)} > l_k^{(1)} h(p^0) + b_k^{(1)} \overset{(*)}{\geq} l_{K_1}^{(1)} h(p^0) + b_{K_1}^{(1)} = h(p^0) \geq \sum_{i\in I} q_i$, where $(*)$ follows from our ordering of the slopes and the intercepts $(l_k^{(1)}, b_k^{(1)})_{k=1}^{K_1}$ as described in (32). Hence, the inequality under consideration indeed holds for the index set I_+. Thus, $(\mathbf{q}, \mathbf{t}^{(1)}) \in M_{h,l}^{\mathrm{ca}}(\mathbf{p})$. The reformulation in (34) now follows as in Theorem 6. $\square$

Proof of Theorem 10. We first note that, by Theorem 8,

$$\rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) = \sup_{\bar{\mathbf{q}}\in M_h^{\mathrm{ca}}(\mathbf{q})} \sum_{i=1}^m -\bar{q}_i u(f(\mathbf{a},\mathbf{x}_i)) - \sup_{\tilde{\mathbf{q}}\in N_{\tilde{h}}^{\mathrm{cv}}(\mathbf{q})} \sum_{i=1}^m \tilde{q}_i u(f(\mathbf{a},\mathbf{x}_i)).$$

Since $\mathcal{U}_t \subset \mathcal{D}_\phi(\mathbf{p}, r)$, we have that for each iteration t, the cutting-plane procedure yields a lower bound c^t on the true optimal objective value (P), which we denote by $\tilde{c}$. By Assumption 1, $\tilde{c}$ is finite. Hence, we may assume that $|c^s - c^t| \leq \frac{1}{2}\epsilon_{\mathrm{tol}}$, for all s, t sufficiently large. Otherwise, the lower bounds will attain or exceed $\tilde{c}$ after finitely many iterations. Assume now that $s > t$. Let $\mathbf{q}^t$ be the worst-case probability vector added to $\mathcal{U}_t$ and let $\mathbf{a}_t, \mathbf{a}_s$ be the optimal solutions obtained at the t, s-th iteration. By definition,

$$\sup_{\bar{\mathbf{q}}\in M_h^{\mathrm{ca}}(\mathbf{q}^t)} \sum_{i=1}^m -\bar{q}_i u(f(\mathbf{a}^t,\mathbf{x}_i)) - \sup_{\tilde{\mathbf{q}}\in N_{\tilde{h}}^{\mathrm{cv}}(\mathbf{q}^t)} \sum_{i=1}^m \tilde{q}_i u(f(\mathbf{a}^t,\mathbf{x}_i)) > c^t + \epsilon_{\mathrm{tol}}, \tag{EC.19}$$

and

$$\sup_{\bar{\mathbf{q}}\in M_h^{\mathrm{ca}}(\mathbf{q}^t)} \sum_{i=1}^m -\bar{q}_i u(f(\mathbf{a}^s,\mathbf{x}_i)) - \sup_{\tilde{\mathbf{q}}\in N_{\tilde{h}}^{\mathrm{cv}}(\mathbf{q}^t)} \sum_{i=1}^m \tilde{q}_i u(f(\mathbf{a}^s,\mathbf{x}_i)) \leq c^s. \tag{EC.20}$$

Therefore,

$$\begin{aligned}
&\sup_{\bar{\mathbf{q}}\in M_h^{\mathrm{ca}}(\mathbf{q}^t)} \sum_{i=1}^m -\bar{q}_i u(f(\mathbf{a}^t,\mathbf{x}_i)) - \sup_{\bar{\mathbf{q}}\in M_h^{\mathrm{ca}}(\mathbf{q}^t)} \sum_{i=1}^m -\bar{q}_i u(f(\mathbf{a}^s,\mathbf{x}_i)) \\
&\qquad + \sup_{\tilde{\mathbf{q}}\in N_{\tilde{h}}^{\mathrm{cv}}(\mathbf{q}^t)} \sum_{i=1}^m \tilde{q}_i u(f(\mathbf{a}^s,\mathbf{x}_i)) - \sup_{\tilde{\mathbf{q}}\in N_{\tilde{h}}^{\mathrm{cv}}(\mathbf{q}^t)} \sum_{i=1}^m \tilde{q}_i u(f(\mathbf{a}^t,\mathbf{x}_i)) \\
&> \epsilon_{\mathrm{tol}} - (c^s - c^t) \geq \frac{1}{2}\epsilon_{\mathrm{tol}}.
\end{aligned}$$

We can further upper bound the differences of the suprema as follows:

$$\begin{aligned}
&\sup_{\bar{\mathbf{q}}\in M_h^{\mathrm{ca}}(\mathbf{q}^t)} \sum_{i=1}^m -\bar{q}_i u(f(\mathbf{a}^t,\mathbf{x}_i)) - \sup_{\bar{\mathbf{q}}\in M_h^{\mathrm{ca}}(\mathbf{q}^t)} \sum_{i=1}^m -\bar{q}_i u(f(\mathbf{a}^s,\mathbf{x}_i)) \\
&\leq \sup_{\bar{\mathbf{q}}\in M_h^{\mathrm{ca}}(\mathbf{q}^t)} \left| \sum_{i=1}^m -\bar{q}_i (u(f(\mathbf{a}^s,\mathbf{x}_i)) - u(f(\mathbf{a}^t,\mathbf{x}_i))) \right| \\
&\leq \sup_{\bar{\mathbf{q}}\in M_h^{\mathrm{ca}}(\mathbf{q}^t)} \|\bar{\mathbf{q}}\|_2 \|(u(f(\mathbf{a}^s,\mathbf{x}_i)) - u(f(\mathbf{a}^t,\mathbf{x}_i)))_{i=1}^m\|_2 \\
&\leq \|(u(f(\mathbf{a}^s,\mathbf{x}_i)) - u(f(\mathbf{a}^t,\mathbf{x}_i)))_{i=1}^m\|_2.
\end{aligned}$$

Similarly,

$$\sup_{\tilde{\mathbf{q}}\in N_{\tilde{h}}^{\mathrm{cv}}(\mathbf{q}^t)} \sum_{i=1}^m \tilde{q}_i u(f(\mathbf{a}^s,\mathbf{x}_i)) - \sup_{\tilde{\mathbf{q}}\in N_{\tilde{h}}^{\mathrm{cv}}(\mathbf{q}^t)} \sum_{i=1}^m \tilde{q}_i u(f(\mathbf{a}^t,\mathbf{x}_i)) \leq \|(u(f(\mathbf{a}^s,\mathbf{x}_i)) - u(f(\mathbf{a}^t,\mathbf{x}_i)))_{i=1}^m\|_2.$$

Therefore,

$$\|(u(f(\mathbf{a}^s, \mathbf{x}_i)) - u(f(\mathbf{a}^t, \mathbf{x}_i)))_{i=1}^m\|_2 \geq \frac{1}{4}\epsilon_{\text{tol}}.$$

However, we have the assumption that $\sup_{\mathbf{a}\in\mathcal{A}, i\in[m]} |u(f(\mathbf{a}, \mathbf{x}_i))| < \infty$. Hence, this leads to a similar contradiction as in the proof of Theorem 3. Therefore, the cutting-plane procedure must terminate. $\square$

Proof of Theorem 11. Let $M \triangleq \sup_{\mathbf{a}\in\mathcal{A}, i\in[m]} |u(f(\mathbf{a}, \mathbf{x}_i))| < \infty$. Similar to the proof of Theorem 7, for any two distortion functions h_1, h_2 such that $\sup_{p\in[0,1]} |h_1(p) - h_2(p)| \leq \epsilon$, we have that

$$\begin{aligned}\rho_{u,h_1,\mathbf{p}}(f(\mathbf{a}, \mathbf{X})) &\leq \rho_{u,h_2,\mathbf{p}}(f(\mathbf{a}, \mathbf{X})) + \epsilon \cdot \left(\max_{i\in[m]} u(f(\mathbf{a}, \mathbf{x}_i)) - \min_{i\in[m]} u(f(\mathbf{a}, \mathbf{x}_i))\right)\\ &\leq \rho_{u,h_2,\mathbf{p}}(f(\mathbf{a}, \mathbf{X})) + 2M\epsilon.\end{aligned}$$

Since $2M\epsilon$ does not depend on both $\mathbf{a}$ and $\mathbf{q}$, we have that, in the nominal case,

$$\left|\min_{\mathbf{a}\in\mathcal{A}} \rho_{u,h_1,\mathbf{p}}(f(\mathbf{a}, \mathbf{X})) - \min_{\mathbf{a}\in\mathcal{A}} \rho_{u,h_2,\mathbf{p}}(f(\mathbf{a}, \mathbf{X}))\right| \leq 2M\epsilon,$$

and similarly, in the robust case,

$$\left|\min_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_1,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) - \min_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_2,\mathbf{q}}(f(\mathbf{a}, \mathbf{X}))\right| \leq 2M\epsilon,$$

which both approach zero as $\epsilon \to 0$. Since Algorithm 3 yields a final objective value c^* such that

$$\left|c^* - \min_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_\epsilon,\mathbf{q}}(f(\mathbf{a}, \mathbf{X}))\right| \leq \epsilon_{\text{tol}},$$

it follows that

$$\left|c^* - \min_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a}, \mathbf{X}))\right| \leq \epsilon_{\text{tol}} + \epsilon 2M \to 0,$$

as $\epsilon_{\text{tol}}, \epsilon \to 0$. $\square$

Proof of Theorem 12. The proof that Algorithm 3 terminates after finitely many steps for problem (P-constraint) is similar to the proof of Theorem 10, where we now take $c^s = c^t = c$, with c the constraint value in (P-constraint). To show convergence, we note that $\mathcal{U}_j \subset \mathcal{D}_\phi(\mathbf{p}, r)$. Hence, $g(\mathbf{a}_{\epsilon_{\text{tol}}})$ is always a lower bound of (P-constraint), for which we denote its optimal objective value by $P(0)$. We define

$$P(\epsilon_{\text{tol}}) \triangleq \min_{\mathbf{a}\in\mathcal{A}} \left\{ g(\mathbf{a}) \;\middle|\; \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) \leq c + \epsilon_{\text{tol}} \right\}.$$

Then, by Algorithm 3, the final solution $\mathbf{a}_{\epsilon_{\text{tol}}}$ is feasible for the minimization problem of $P(\epsilon_{\text{tol}})$. Hence, $P(\epsilon_{\text{tol}}) \leq g(\mathbf{a}_{\epsilon_{\text{tol}}}) \leq P(0)$. It remains to show that $\lim_{\epsilon_{\text{tol}}\to 0} P(\epsilon_{\text{tol}}) = P(0)$, which follows from the proof of Theorem 13 below. $\square$

Proof of Theorem 13. Let

$$L_\epsilon \triangleq \min_{\mathbf{a}\in\mathcal{A}} \left\{ g(\mathbf{a}) \;\middle|\; \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_\epsilon,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \leq c \right\}.$$

Then, by the proof of Theorem 11, for all $\mathbf{a}\in\mathcal{A}$,

$$\left| \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h_\epsilon,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) - \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \right| \leq \epsilon M,$$

where $M \triangleq 2\sup_{\mathbf{a}\in\mathcal{A}, i\in[m]} |u(f(\mathbf{a},\mathbf{x}_i))| < \infty$, which follows from the continuity of $u(f(\mathbf{a},\mathbf{x}_i))$ in $\mathbf{a}$ and the compactness of $\mathcal{A}$. Hence, we conclude that $P(\epsilon) \leq L_\epsilon \leq P(0)$, where $P(0)$ is the optimal objective value of (P-constraint), and

$$P(\epsilon) \triangleq \min_{\mathbf{a}\in\mathcal{A}} \left\{ g(\mathbf{a}) \;\middle|\; \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \leq c + \epsilon M \right\}.$$

We will now show that $\lim_{\epsilon\downarrow 0} P(\epsilon) = P(0)$, hence concluding the proof. To show this, we use Berge's maximum theorem (Berge, 1963), for which we have to show that the set-valued function

$$\epsilon \mapsto G(\epsilon) \triangleq \left\{ \mathbf{a}\in\mathcal{A} : \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) \leq c + \epsilon M \right\}$$

is a compact-valued, continuous correspondence at $\epsilon = 0$. We first examine compactness, which entails that for each $\epsilon \geq 0$, the set $G(\epsilon)$ must be compact. We note that since $\mathcal{A}$ is compact, we have that $G(\epsilon)$ is bounded for all $\epsilon \geq 0$. Hence, we only need to show that $G(\epsilon)$ is closed. This holds if $\mathbf{a} \mapsto \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X}))$ is continuous, which can be proven as follows: by Lemma EC.2, we have that $(\mathbf{q},\mathbf{a}) \mapsto \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X}))$ is jointly continuous in $(\mathbf{q},\mathbf{a})$. Since the set $\mathcal{D}_\phi(\mathbf{p},r)$ is compact and independent of $\mathbf{a}$, Berge's maximum theorem implies that $\mathbf{a} \mapsto \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X}))$ is continuous. Hence, $G(\epsilon)$ is compact.

We now show that $G(\epsilon)$ is lower-and upper-hemicontinuous at $\epsilon = 0$. For upper-hemicontinuity, we must show that $G(0)$ is non-empty, which holds due to Assumption 4, and that, for any $\epsilon_j \to 0$ and any sequence $\mathbf{a}_j \in G(\epsilon_j)$, there exists a convergent subsequence $\mathbf{a}_{j_k}$ such that $\mathbf{a}_{j_k} \to \mathbf{a}_0 \in G(0)$. Note that due to the compactness of $\mathcal{A}$, there is indeed a subsequence $\mathbf{a}_{j_k} \to \mathbf{a}_0 \in \mathcal{A}$. It remains to show that we have $\mathbf{a}_0 \in G(0)$. This follows from continuity: $\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a}_0,\mathbf{X})) = \lim_{k\to\infty} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a}_{j_k},\mathbf{X})) \leq \lim_{k\to\infty} c + \epsilon_{j_k} M = c$.

Finally, for lower-hemicontinuity at $\epsilon = 0$, we have to show that, for every $\mathbf{a}_0 \in G(0)$ and every sequence $\epsilon_j \to 0$, there exist a $J \geq 1$ and a sequence $\mathbf{a}_j$ such that, for all $j \geq J$, we have $\mathbf{a}_j \in G(\epsilon_j)$ and $\mathbf{a}_j \to \mathbf{a}_0$. However, since $\epsilon_j > 0$ for all j, we can take the sequence $\mathbf{a}_j = \mathbf{a}_0 \in G(\epsilon_j)$. Hence, we may apply Berge's maximum theorem to conclude that $\lim_{\epsilon\downarrow 0} P(\epsilon) = P(0)$, which concludes the proof. $\square$

Lemma EC.2. *Let $u(f(\mathbf{a},\mathbf{x}_i))$ be continuous in $\mathbf{a}$ for all $i\in[m]$ and let h be a continuous distortion function. Then, the rank-dependent evaluation function*

$$(\mathbf{q},\mathbf{a}) \mapsto \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})),$$

is jointly continuous in $(\mathbf{q},\mathbf{a})$.

Proof. For any $\mathbf{a} \in \mathcal{A}$, we denote the indices $1(\mathbf{a}), \dots, m(\mathbf{a})$ such that

$$-u(f(\mathbf{a}, \mathbf{x}_{1(\mathbf{a})})) \leq \cdots \leq -u(f(\mathbf{a}, \mathbf{x}_{m(\mathbf{a})})),$$

and we define

$$\Delta u(f(\mathbf{a}, \mathbf{x}_{1(\mathbf{a})})) \triangleq -u(f(\mathbf{a}, \mathbf{x}_{1(\mathbf{a})})),\ \Delta u(f(\mathbf{a}, \mathbf{x}_{i(\mathbf{a})})) \triangleq u(f(\mathbf{a}, \mathbf{x}_{(i-1)(\mathbf{a})})) - u(f(\mathbf{a}, \mathbf{x}_{i(\mathbf{a})})),$$

for $i = 2, \dots, m$. Denote the index set

$$\mathcal{I}_+(\mathbf{a}_0) = \{i \in \{2, \dots, m\} : \Delta u(f(\mathbf{a}_0, \mathbf{x}_{i(\mathbf{a}_0)})) > 0\}.$$

We enumerate the elements in $\mathcal{I}_+(\mathbf{a}_0)$ as $i_1 < \cdots < i_K$, for $K = |\mathcal{I}_+(\mathbf{a}_0)| \leq m-1$. Set $i_0 = 1$ and $i_{K+1} = m+1$. Then, the index set $[m]$ is partitioned into $K+1$ disjoint classes $\Pi_j \triangleq \{i(\mathbf{a}_0) : i_{j-1} \leq i < i_j\}$, with $j \in [K+1]$, that have the following two properties: (i) If $k, l \in [m]$ are indices of adjacent classes, e.g., $k \in \Pi_j$ and $l \in \Pi_{j+1}$ for some j, then $u(f(\mathbf{a}_0, \mathbf{x}_k)) - u(f(\mathbf{a}_0, \mathbf{x}_l)) = \Delta u(f(\mathbf{a}_0, \mathbf{x}_{i_{j+1}(\mathbf{a}_0)})) > 0$. (ii) If k, l belong to the same class Π_j, we have $u(f(\mathbf{a}_0, \mathbf{x}_k)) - u(f(\mathbf{a}_0, \mathbf{x}_l)) = 0$. Continuity of $u(f(\mathbf{a}, \mathbf{x}))$ in $\mathbf{a}_0$ implies that for all $\mathbf{a}$ sufficiently close to $\mathbf{a}_0$, the ranking indices $1(\mathbf{a}), \dots, m(\mathbf{a})$ can also be partitioned such that $\{i(\mathbf{a}) : i_{j-1} \leq i < i_j\} = \{i(\mathbf{a}_0) : i_{j-1} \leq i < i_j\}$, for all $j \in [K+1]$. In particular, this means that within each class, the ranking indices of $\mathbf{a}$ constitute a permutation of that of $\mathbf{a}_0$.

Hence, we have

$$\begin{aligned}
&\rho_{u,h,\mathbf{q}}(f(\mathbf{a}, \mathbf{X})) \\
&= \sum_{i=1}^{m} h\left(\sum_{k=i}^{m} q_{k(\mathbf{a})}\right) \Delta u(f(\mathbf{a}, \mathbf{x}_{i(\mathbf{a})})) = \sum_{j=1}^{K+1} \sum_{i=i_{j-1}}^{i_j - 1} h\left(\sum_{k=i}^{m} q_{k(\mathbf{a})}\right) \Delta u(f(\mathbf{a}, \mathbf{x}_{i(\mathbf{a})})) \\
&= \sum_{j=1}^{K+1} h\left(\sum_{k=i_{j-1}}^{m} q_{k(\mathbf{a})}\right) \Delta u(f(\mathbf{a}, \mathbf{x}_{i_{j-1}(\mathbf{a})})) + \sum_{j=1}^{K+1} \sum_{i=i_{j-1}+1}^{i_j - 1} h\left(\sum_{k=i}^{m} q_{k(\mathbf{a})}\right) \Delta u(f(\mathbf{a}, \mathbf{x}_{i(\mathbf{a})})) \\
&\overset{(\mathbf{q},\mathbf{a})\to(\mathbf{q}_0,\mathbf{a}_0)}{=} \sum_{j=1}^{K+1} h\left(\sum_{k=i_{j-1}}^{m} q_{0,k(\mathbf{a}_0)}\right) \Delta u(f(\mathbf{a}_0, \mathbf{x}_{i_{j-1}(\mathbf{a}_0)})) + 0 \\
&= \rho_{u,h,\mathbf{q}_0}(f(\mathbf{a}_0, \mathbf{X})).
\end{aligned}$$

This completes the proof. □

EC.3 Derivations of the Conjugate Functions and Their Epigraphs

In this appendix, we provide detailed derivations of the explicit conjugate functions, as well as the epigraphs of $\lambda(-h)^*(\frac{-\nu}{\lambda}) \leq z, \lambda > 0$, and $\gamma\phi^*(\frac{s}{\gamma}) \leq t, \gamma > 0$, for the collection of canonical examples in Tables EC.1–EC.2; we only provide the details in case the derivations are considered to be non-trivial.

EC.3.1 Distortion Functions

- For $h(p) = \begin{cases} (1+r)p & p < 1/2 \\ (1-r)p + r & p \geq 1/2 \end{cases}$ and $0 < r < 1$, we examine the conjugate

$$(-h)^*(y) = \max\{\sup_{t \in [0, \frac{1}{2})} (y + (1+r))t, \sup_{t \geq \frac{1}{2}} (y + (1-r))t + r\}.$$

We have $(-h)^*(y) = \infty$ for $y > -(1-r)$. For $-(1+r) \le y \le -(1-r)$, we have $(-h)^*(y) = \max\{1/2(y+(1+r)), 1/2(y+(1-r))+r\} = 1/2(y+(1+r))$. For $y < -(1+r)$, $(-h)^*(y) = \max\{0, 1/2(y+(1-r))+r\} = 0$. Therefore, $(-h)^*(y) = \max\{(y+1+r)/2, 0\}$, for $y \le -(1-r)$.

- Let $h(p) = (1+r)p - rp^2$, $0 < r < 1$. For $y \le 0$, we have

$$(-h)^*(y) = \sup_{t\in[0,1]} \{ty + (1+r)t - rt^2\} = \sup_{t\in[0,1]} \{(y+(1+r))t - rt^2\}.$$

Differentiating the objective w.r.t. t yields

$$y + (1+r) - 2rt,$$

which is non-negative for $t \le \frac{y+1+r}{2r}$ and negative otherwise. If $\frac{y+1+r}{2r} < 0$, then the derivative is always negative, hence the maximum is obtained at $t = 0$ and thus we have $(-h)^*(y) = 0$. If $\frac{y+1+r}{2r} > 1$, then the derivative is always positive, hence $(-h)^*(y) = y+1$. If $\frac{y+1+r}{2r} \in [0,1]$, then the maximum is attained at $t = \frac{y+1+r}{2r}$. Hence, for $y \le 0$, we have

$$(-h)^*(y) = \begin{cases} \frac{1}{4r}\max\{y+1+r, 0\}^2 & y+1 \le r \\ y+1 & y+1 > r. \end{cases}$$

The epigraph of $(-h)^*$ can be represented by

$$\begin{cases} t = t_1 + t_2 + r \\ y + 1 - r = y_1 + y_2 \\ t_1 \ge \frac{1}{4r}\max\{y_1 + 2r, 0\}^2 - r, \;\; t_2 \ge y_2 \\ y_1 \le 0, y_2 \ge 0. \end{cases}$$

Indeed, let $(y,t) \in \mathrm{Epi}((-h)^*)$. If $y+1-r \le 0$, then we can choose $y_1 = y+1-r$, $t_1 = t-r$, $y_2 = t_2 = 0$. If $y+1-r \ge 0$, then choose $y_1 = t_1 = 0$, $y_2 = y+1-r$, $t_2 = t-r$.

Conversely, let $(y, t, y_1, y_2, t_1, t_2)$ satisfy the above constraints. Define

$$f(z) = \begin{cases} \frac{1}{4r}\max\{z+2r, 0\}^2 - r & z \le 0 \\ z & z \ge 0. \end{cases}$$

Then, $(-h)^*(y) = f(y+1-r)+r$. Hence, $(-h)^*(y) \le t \Leftrightarrow f(y+1-r) \le t-r$. Since $y_1 \le 0$, $y_2 \ge 0$, we have that $y_1 \le y_1 + y_2 \le y_2$. Since f is convex and $f(0) = 0$, we have

$$\frac{f(y_1+y_2) - f(y_1)}{y_2} \le \frac{f(y_2) - f(0)}{y_2},$$

and thus $f(y_1+y_2) \le f(y_1) + f(y_2)$. Therefore, $(-h)^*(y) = f(y+1-r)+r = f(y_1+y_2) \le f(y_1) + f(y_2) + r \le t_1 + t_2 + r = t$.

The epigraph of $\lambda(-h)^*(\frac{-\nu}{\lambda}) \le z$ is then given by

$$\begin{cases} z = z_1 + z_2 \\ -\nu + \lambda(1-r) = \xi_1 + \xi_2 \\ (z_1 + \lambda) \ge \sqrt{\frac{w^2}{r} + (z_1 - \lambda)^2}, \;\; z_2 \ge \xi_2 \\ \xi_1 + 2r \le w \\ \xi_1 \le 0, \xi_2, w, \nu \ge 0. \end{cases}$$

- For $h(p) = \begin{cases} 1-(1-p)^n & 0 \le p < 1 \\ 1 & p \ge 1 \end{cases}$, we will derive a tractable reformulation of the epigraph $(-h)^*$ using duality. We have

$$\mathrm{Epi}(-h) = \{(p,t) \in \mathbb{R}_{\ge 0} \times \mathbb{R} : \exists\ (u_1,u_2) \in \mathbb{R}^2_{\ge 0} : u_1 - 1 \le t, u_2 \le u_1^{1/n}, 1-p \le u_2\}.$$

Indeed, let $(p,t) \in \mathrm{Epi}(-h)$. If $0 \le p < 1$, then $(1-p)^n - 1 \le t$. We can choose $u_2 = 1-p \ge 0$ and $u_1 = u_2^n \ge 0$. If $p \ge 1$, then $-1 \le t$. We can choose $u_1 = u_2 = 0$. Conversely, let (p,t,u_1,u_2) satisfy the above constraints. If $p \ge 1$, we have $t \ge -1 + u_1 \ge -1$. Thus, $(p,t) \in \mathrm{Epi}(-h)$. If $0 \le p < 1$, then $(1-p)^n - 1 \le u_2^n - 1 \le u_1 - 1 \le t$. Hence, we also have $(p,t) \in \mathrm{Epi}(-h)$.

Consider the epigraph of $(-h)^*$. We have

$$\mathrm{Epi}((-h)^*) = \{(y,s) : yp - (-h)(p) \le s, \forall p \ge 0\} = \{(y,s) : yp - t \le s, \forall (p,t) \in \mathrm{Epi}(-h)\}.$$

Therefore, we have that $(y,s) \in \mathrm{Epi}((-h)^*)$ if and only if the optimization problem

$$\min_{p,u_1,u_2 \ge 0, t \in \mathbb{R}} \{-yp + t | u_1 - 1 \le t, u_2 \le u_1^{1/n}, 1-p \le u_2\}$$

is bounded below by $-s$. Since this is a convex problem with a point $p = 1, u_2 = 0, u_1 = 1, t = 0$ that satisfies Slater's condition, we may apply the duality theorem and obtain that this optimization problem is equal to

$$\max_{\xi_1,\xi_2,\xi_3 \ge 0} \inf_{p,u_1,u_2 \ge 0, t \in \mathbb{R}} -yp + t + \xi_1(u_1 - 1 - t) + \xi_2(u_2 - u_1^{1/n}) + \xi_3(1 - p - u_2),$$

which, after some rewriting is equal to

$$\max_{\xi_2,\xi_3 \ge 0} \{-1 + \xi_3 + (n^{-\frac{n}{n-1}} - n^{-\frac{1}{n-1}})\xi_2^{\frac{n}{n-1}} \mid y + \xi_3 \le 0, \xi_2 \ge \xi_3\}.$$

Therefore, we have

$$\begin{aligned}
&\mathrm{Epi}((-h)^*) \\
&= \{(y,s) : \exists\ \xi_2,\xi_3 \ge 0 : 1 - \xi_3 + (n^{-\frac{1}{n-1}} - n^{-\frac{n}{n-1}})\xi_2^{\frac{n}{n-1}} \le s, y + \xi_3 \le 0, \xi_2 \ge \xi_3\} \\
&= \{(y,s) : \exists\ \xi_2,\xi_3,\xi_4 \ge 0 : 1 - \xi_3 + (n^{-\frac{1}{n-1}} - n^{-\frac{n}{n-1}})\xi_4 \le s, \xi_2 \le \xi_4^{\frac{n-1}{n}} \cdot 1^{1-\frac{n-1}{n}}, \\
&\qquad y + \xi_3 \le 0, \xi_2 \ge \xi_3\}.
\end{aligned}$$

This gives the reformulation of the epigraph of the perspective $\lambda(-h)^*(\frac{-\nu}{\lambda}) \le z$ as

$$\lambda(-h)^*(\frac{-\nu}{\lambda}) \le z \Leftrightarrow \begin{cases} \lambda - \xi_3 + (n^{-\frac{1}{n-1}} - n^{-\frac{n}{n-1}})\xi_4 \le z \\ \xi_2 \le \xi_4^{\frac{n-1}{n}} \cdot \lambda^{1-\frac{n-1}{n}} \\ -\nu + \xi_3 \le 0 \\ \xi_2 \ge \xi_3 \\ \xi_2,\xi_3,\xi_4 \ge 0. \end{cases}$$

- For $h(p) = \begin{cases}(1-(1-p)^n)^{1/n} & 0 \le p < 1\\ 1 & p \ge 1\end{cases}$, we will derive a tractable reformulation of the epigraph $(-h)^*$ using duality. We have

  $$\mathrm{Epi}(-h) = \{(p,t) \in \mathbb{R}_{\ge 0} \times \mathbb{R} : \exists\, u_1, u_2 \ge 0 : u_1 \ge -t, u_2^{1/n} \ge u_1, (1-u_2)^{1/n} \ge 1-p, u_2 \le 1\}.$$

  Indeed, if $(p,t) \in \mathrm{Epi}(-h)$, then for $0 \le p < 1$, we have $-(1-(1-p)^n)^{1/n} \le t$. Choose $u_2 = (1-(1-p)^n)$ and $u_1 = u_2^{1/n}$ gives the right inclusion. If $p \ge 1$, then choose $u_2 = 1 = u_1$, which also gives the right inclusion. Conversely, let (p,t,u_1,u_2) satisfy the above constraints. If $0 \le p < 1$, then, by construction, $(p,t) \in \mathrm{Epi}(-h)$. If $p \ge 1$, we have $u_1 \le u_2^{1/n} \le 1$. Hence, $t \ge -u_1 \ge -1$, thus $(p,t) \in \mathrm{Epi}(-h)$.

  Consider the epigraph of $(-h)^*$. Again, we have

  $$\mathrm{Epi}((-h)^*) = \{(y,s) : yp - t \le s, \forall (p,t) \in \mathrm{Epi}(-h)\}.$$

  Therefore, we have that $(y,s) \in \mathrm{Epi}((-h)^*)$ if and only if the optimization problem

  $$\min_{p,u_1,\ge 0, u_2 \in [0,1], t \in \mathbb{R}} \{-yp + t | u_1 \ge -t, u_2^{1/n} \ge u_1, (1-u_2)^{1/n} \ge 1-p\}$$

  is bounded below by $-s$. Since this is a convex problem with a point $p = 2, u_2 = 1, u_1 = 1/2, t = 1$ that satisfies Slater's condition, we may apply the duality theorem and obtain that this optimization problem is equal to

  $$\max_{\xi_1,\xi_2,\xi_3 \ge 0} \{\xi_3 + \inf_{p,u_1 \ge 0, u_2 \in [0,1], t \in \mathbb{R}} -(y+\xi_3)p + (1-\xi_1)t + (\xi_2 - \xi_1)u_1 - \xi_2 u_2^{1/n} - \xi_3(1-u_2)^{1/n}\}$$
  $$= \max_{\xi_2,\xi_3 \ge 0} \{\xi_3 + \inf_{u_2 \in [0,1]} -\xi_2 u_2^{1/n} - \xi_3(1-u_2)^{1/n} | \xi_1 = 1, y + \xi_3 \le 0, \xi_2 \ge \xi_1\}.$$

  The convex function $-\xi_2 u_2^{1/n} - \xi_3(1-u_2)^{1/n}$ has derivative

  $$\frac{\mathrm{d}}{\mathrm{d}u_2} = \frac{1}{n}(\xi_3(1-u_2)^{\frac{1-n}{n}} - \xi_2 u_2^{\frac{1-n}{n}}),$$

  which has a root at

  $$\begin{aligned}\frac{\xi_2}{u_2^{\frac{n-1}{n}}} &= \frac{\xi_3}{(1-u_2)^{\frac{n-1}{n}}}\\ \xi_2(1-u_2)^{\frac{n-1}{n}} &= \xi_3 u_2^{\frac{n-1}{n}}\\ u_2 &= \frac{\xi_2^{\frac{n}{n-1}}}{\xi_2^{\frac{n}{n-1}} + \xi_3^{\frac{n}{n-1}}} \in [0,1].\end{aligned}$$

  Since we are examining a convex function, this is where the minimum is attained. Hence, we have

  $$\begin{aligned}\inf_{u_2 \in [0,1]} -\xi_2 u_2^{1/n} - \xi_3(1-u_2)^{1/n} &= -\xi_2 \cdot \left(\frac{\xi_2^{\frac{n}{n-1}}}{\xi_2^{\frac{n}{n-1}} + \xi_3^{\frac{n}{n-1}}}\right)^{1/n} - \xi_3 \cdot \left(\frac{\xi_3^{\frac{n}{n-1}}}{\xi_2^{\frac{n}{n-1}} + \xi_3^{\frac{n}{n-1}}}\right)^{1/n}\\ &= -\frac{\xi_2^{\frac{n}{n-1}} + \xi_3^{\frac{n}{n-1}}}{(\xi_2^{\frac{n}{n-1}} + \xi_3^{\frac{n}{n-1}})^{1/n}} = -(\xi_2^{\frac{n}{n-1}} + \xi_3^{\frac{n}{n-1}})^{\frac{n-1}{n}}.\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Epi}((-h)^*) &= \{(y,s) : \exists\ \xi_2, \xi_3 \geq 0 : y + \xi_3 \leq 0, \xi_2 \geq 1, -\xi_3 + (\xi_2^{\frac{n}{n-1}} + \xi_3^{\frac{n}{n-1}})^{\frac{n-1}{n}} \leq s\} \\ &= \{(y,s) : \exists\ \xi_2, \xi_3 \geq 0 : y + \xi_3 \leq 0, \xi_2 \geq 1, -\xi_3 + \|(\xi_2, \xi_3)\|_{\frac{n}{n-1}} \leq s\},\end{aligned}$$

where $\|.\|_p$ is the p-norm for $p \geq 1$. Thus, the reformulation of the epigraph of the perspective $\lambda(-h)^*(\frac{-\nu}{\lambda}) \leq z$ is given by

$$\lambda(-h)^*(\frac{-\nu}{\lambda}) \leq z \Leftrightarrow -\nu + \xi_3 \leq 0,\ \xi_2 \geq \lambda,\ -\xi_3 + \|(\xi_2, \xi_3)\|_{\frac{n}{n-1}} \leq z,\ \xi_2, \xi_3 \geq 0.$$

EC.3.2 Divergence Functions

- For $\phi^*(y) = e^y - 1$, we have

$$\gamma(e^{\frac{s}{\gamma}} - 1) \leq t \Leftrightarrow w - \gamma \leq t,\ e^{\frac{s}{\gamma}} \leq \frac{w}{\gamma} \Leftrightarrow w - \gamma \leq t, \gamma \log(\frac{\gamma}{w}) + s \leq 0.$$

  Note that $\gamma \log(\frac{\gamma}{w})$ is the relative entropy.

- For $\phi^*(y) = 2 - 2\sqrt{1-y}$, $y < 1$, we have

$$\gamma(2 - 2\sqrt{1 - \frac{s}{\gamma}}) \leq t, s < \gamma \Leftrightarrow 2\gamma - 2\gamma\sqrt{\frac{\gamma - s}{\gamma}} \leq t, s < \gamma$$

$$\begin{aligned}\Leftrightarrow 2\gamma - 2\gamma\sqrt{\frac{v}{\gamma}} \leq t, v = \gamma - s, s < \gamma &\Leftrightarrow 2\gamma - 2\sqrt{\gamma v} \leq t, v = \gamma - s, v > 0 \\ &\Leftrightarrow 2\gamma - 2w \leq t, w^2 \leq \gamma v, w \geq 0, v = \gamma - s, v > 0 \\ &\Leftrightarrow 2\gamma - 2w \leq t, \sqrt{w^2 + \frac{1}{4}(\gamma - v)^2} \leq \frac{1}{2}(\gamma + v), w \geq 0, \\ &\qquad v = \gamma - s, v > 0,\end{aligned}$$

  where in the last equivalence we used the equality $xy = \frac{1}{4}((x+y)^2 - (x-y)^2)$.

- For $\phi^*(y) = \frac{y}{1-y} = -1 + \frac{1}{1-y}, y < 1$, we have

$$\begin{aligned}-1 + \frac{1}{1-y} \leq t, y < 1 &\Leftrightarrow -1 + v \leq t, \frac{1}{w} \leq v, w = 1 - y, w > 0 \\ &\Leftrightarrow -1 + v \leq t, \sqrt{1 + \frac{1}{4}(v - w)^2} \leq \frac{1}{2}(v + w), w = 1 - y, w > 0.\end{aligned}$$

  The epigraph of the perspective function can then be obtained as

$$\gamma\phi^*(\frac{s}{\gamma}) \leq t \Leftrightarrow -\gamma + v \leq t, \sqrt{\gamma^2 + \frac{1}{4}(v - w)^2} \leq \frac{1}{2}(v + w), w = \gamma - s, w > 0.$$

- For $\phi^*(y) = y + (\theta - 1)(\frac{|y|}{\theta})^{\frac{\theta}{\theta-1}}$. We show that the epigraph can be represented by a power cone, which is $\mathcal{P}_3^{\alpha,1-\alpha} \triangleq \{x \in \mathbb{R}^3 : x_1^\alpha x_2^{1-\alpha} \geq |x_3|, x_1, x_2 \geq 0\}$, for $0 < \alpha < 1$. More on tractability of power cones can be found in Chares (2009). We have that

$$\begin{aligned}y + (\theta - 1)(\frac{|y|}{\theta})^{\frac{\theta}{\theta-1}} \leq t &\Leftrightarrow y + (\theta - 1)\theta^{\frac{\theta}{1-\theta}} w \leq t,\ |y|^{\frac{\theta}{\theta-1}} \leq w \\ &\Leftrightarrow y + (\theta - 1)\theta^{\frac{\theta}{1-\theta}} w \leq t,\ |y| \leq w^{\frac{\theta}{\theta-1}} \cdot 1^{1-\frac{\theta}{\theta-1}},\end{aligned}$$

is conic quadratic representable. Hence, so is the epigraph of the perspective

$$\gamma\phi^*(\frac{s}{\gamma}) \leq t \Leftrightarrow s + (\theta-1)\theta^{\frac{\theta}{1-\theta}} w \leq t,\ |s| \leq w^{\frac{\theta}{\theta-1}} \cdot \gamma^{1-\frac{\theta}{\theta-1}}.$$

- Similarly, for $\phi^*(y) = \frac{1}{\theta}(1-y(1-\theta))^{\frac{\theta}{\theta-1}} - \frac{1}{\theta}, y < \frac{1}{1-\theta}$, we have that

$$\frac{1}{\theta}(1-y(1-\theta))^{\frac{\theta}{\theta-1}} - \frac{1}{\theta} \leq t, y < \frac{1}{1-\theta} \Leftrightarrow |w| \leq (t\theta+1)^{\frac{\theta-1}{\theta}} \cdot 1^{1-\frac{\theta-1}{\theta}}, w = 1-y(1-\theta),$$
$$y < \frac{1}{1-\theta},$$

is conic quadratic. Hence, so is the epigraph of the perspective

$$\gamma\phi^*(\frac{s}{\gamma}) \leq t \Leftrightarrow |w| \leq (t\theta+\gamma)^{\frac{\theta-1}{\theta}} \cdot \gamma^{1-\frac{\theta-1}{\theta}}, w = \gamma - s(1-\theta), s < \frac{\gamma}{1-\theta}.$$

EC.4 Robust Rank-Dependent Evaluation Using Penalization

This section investigates the comparison between

$$\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(\cdot), \tag{EC.21}$$

and

$$\tilde{\rho}_{\text{rob}}(\cdot) \triangleq \sup_{\mathbf{q}\in\Delta_m} \rho_{u,h,\mathbf{q}}(\cdot) - \theta I_\phi(\mathbf{q},\mathbf{p}), \quad \theta > 0, \tag{EC.22}$$

and associated optimization problems, where Δ_m denotes the set of all m-dimensional probability vectors.

Lemma EC.3. *Let $h : [0,1] \to [0,1]$ be a concave distortion function. Then, the robust risk measure $\tilde{\rho}_{rob}$ defined in* (EC.22) *admits the dual representation*

$$\tilde{\rho}_{rob}(X) = \sup_{\bar{\mathbf{q}}\in\Delta_m} \mathbb{E}_{\bar{\mathbf{q}}}[-u(X)] - c(\bar{\mathbf{q}}),$$

with ambiguity index

$$c(\bar{\mathbf{q}}) \triangleq \inf_{\mathbf{q}\in\Delta_m} \theta I_\phi(\mathbf{q},\mathbf{p}) + \alpha(\bar{\mathbf{q}},\mathbf{q}), \qquad \alpha(\bar{\mathbf{q}},\mathbf{q}) \triangleq \begin{cases} 0, & \text{if } \bar{\mathbf{q}} \in M_h(\mathbf{q}); \\ \infty, & \text{else}; \end{cases}$$

where $M_h(\mathbf{q})$ is as defined in (15).

Proof. By Denneberg (1994), Proposition 10.3, we have the dual representation

$$\rho_{u,h,\mathbf{q}}(X) = \sup_{\bar{\mathbf{q}}\in\Delta_m} \mathbb{E}_{\bar{\mathbf{q}}}[-u(X)] - \alpha(\bar{\mathbf{q}},\mathbf{q}).$$

Hence, we derive

$$
\begin{aligned}
\tilde{\rho}_{\text{rob}}(X) &= \sup_{\mathbf{q}\in\Delta_m} \rho_{u,h,\mathbf{q}}(X) - \theta \sum_{i=1}^{m} p_i \phi\left(\frac{q_i}{p_i}\right) \\
&= \sup_{\mathbf{q}\in\Delta_m} \sup_{\bar{\mathbf{q}}\in\Delta_m} \mathbb{E}_{\bar{\mathbf{q}}}[-u(X)] - \alpha(\bar{\mathbf{q}},\mathbf{q}) - \theta \sum_{i=1}^{m} p_i \phi\left(\frac{q_i}{p_i}\right) \\
&= \sup_{\bar{\mathbf{q}}\in\Delta_m} \sup_{\mathbf{q}\in\Delta_m} \mathbb{E}_{\bar{\mathbf{q}}}[-u(X)] - \alpha(\bar{\mathbf{q}},\mathbf{q}) - \theta \sum_{i=1}^{m} p_i \phi\left(\frac{q_i}{p_i}\right) \\
&= \sup_{\bar{\mathbf{q}}\in\Delta_m} \mathbb{E}_{\bar{\mathbf{q}}}[-u(X)] - \inf_{\mathbf{q}\in\Delta_m} \left(\alpha(\bar{\mathbf{q}},\mathbf{q}) + \theta \sum_{i=1}^{m} p_i \phi\left(\frac{q_i}{p_i}\right)\right) \\
&= \sup_{\bar{\mathbf{q}}\in\Delta_m} \mathbb{E}_{\bar{\mathbf{q}}}[-u(X)] - c(\bar{\mathbf{q}}).
\end{aligned}
$$

□

We next show that a decision-maker who minimizes (EC.21) obtains the same minimizer as when minimizing (EC.22), for a specific θ.

Proposition EC.1. *Assume the existence of a minimizer* $\mathbf{a}^* \in \operatorname{argmin}_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X}))$. *Then for each* $r > 0$, *there exists a* θ^*, *such that*

$$
\mathbf{a}^* \in \operatorname{argmin}_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\Delta_m} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) - \theta^* I_\phi(\mathbf{q},\mathbf{p}).
$$

Proof. As we have shown in Lemma EC.3, $\min_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\Delta_m} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X})) - \theta I_\phi(\mathbf{q},\mathbf{p})$ is equivalent to

$$
\min_{\mathbf{a}\in\mathcal{A}} \left\{ \sup_{\substack{\mathbf{q}\geq\mathbf{0}\\ \mathbf{q}^T\mathbf{1}=1}} \sup_{\bar{\mathbf{q}}\in M_h(\mathbf{q})} -\sum_{i=1}^{m} \bar{q}_i u(f(\mathbf{a},\mathbf{x}_i)) - \theta \sum_{i=1}^{m} p_i \phi\left(\frac{q_i}{p_i}\right) \right\}, \tag{EC.23}
$$

for any fixed constant $\theta > 0$. Let $\mathbf{a}^*$ be a solution of $\min_{\mathbf{a}\in\mathcal{A}} \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \rho_{u,h,\mathbf{q}}(f(\mathbf{a},\mathbf{X}))$. Then, strong duality implies the existence of a γ^* (depending on $\mathbf{a}^*$), such that

$$
\begin{aligned}
&\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \sup_{\bar{\mathbf{q}}\in M_h(\mathbf{q})} -\sum_{i=1}^{m} \bar{q}_i u(f(\mathbf{a}^*,\mathbf{x}_i)) \\
&= \gamma^* r + \sup_{\substack{\mathbf{q}\geq\mathbf{0}\\ \mathbf{q}^T\mathbf{1}=1}} \sup_{\bar{\mathbf{q}}\in M_h(\mathbf{q})} -\sum_{i=1}^{m} \bar{q}_i u(f(\mathbf{a}^*,\mathbf{x}_i)) - \gamma^* \sum_{i=1}^{m} p_i \phi\left(\frac{q_i}{p_i}\right).
\end{aligned} \tag{EC.24}
$$

Since $\mathbf{a}^*$ is a minimizer, we also have

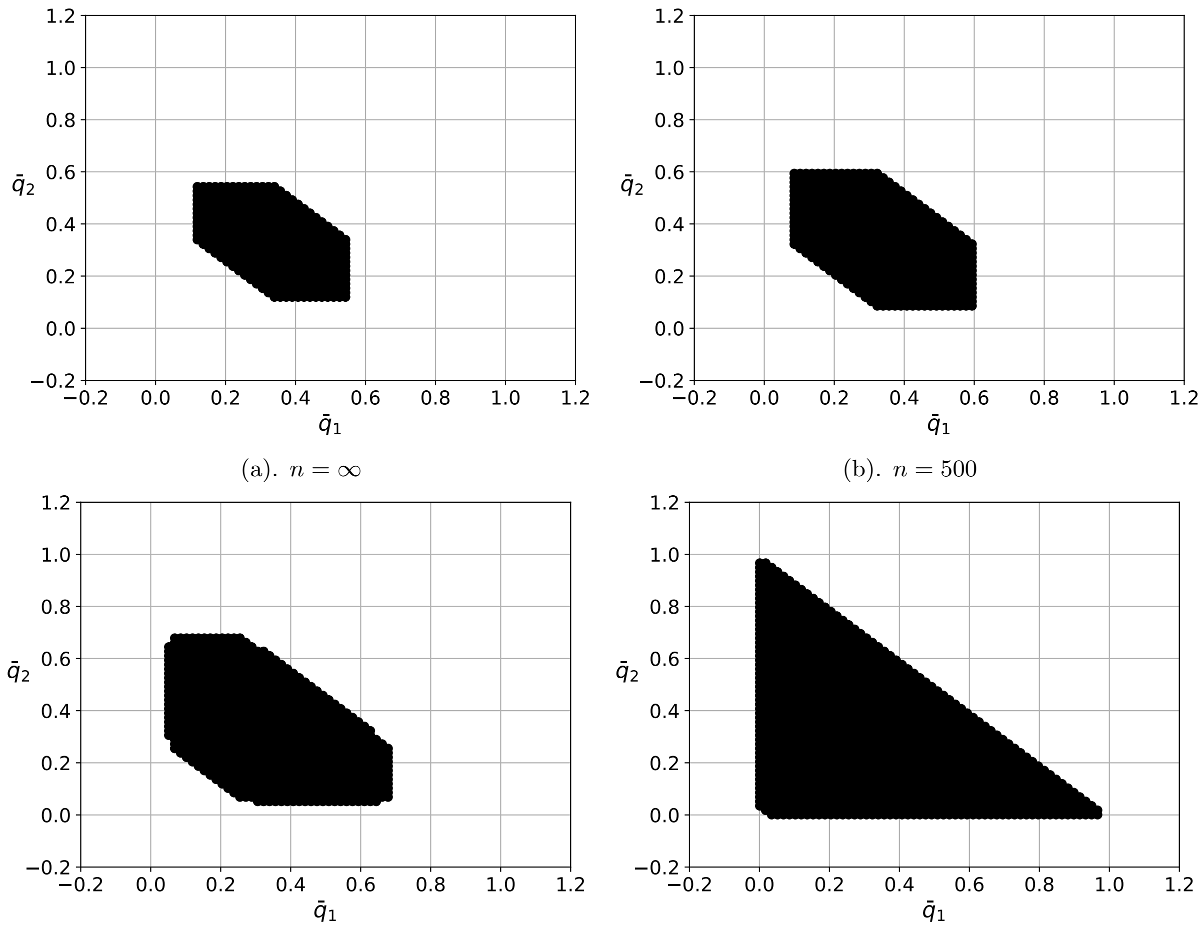
$$\begin{aligned}
&\sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \sup_{\bar{\mathbf{q}}\in M_h(\mathbf{q})} -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}^*,\mathbf{x}_i)) \\
&\leq \sup_{\mathbf{q}\in\mathcal{D}_\phi(\mathbf{p},r)} \sup_{\bar{\mathbf{q}}\in M_h(\mathbf{q})} -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a},\mathbf{x}_i)) \\
&= \inf_{\gamma\geq 0}\left\{\gamma r + \sup_{\substack{\mathbf{q}\geq\mathbf{0}\\ \mathbf{q}^T\mathbf{1}=1}} \sup_{\bar{\mathbf{q}}\in M_h(\mathbf{q})} -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a},\mathbf{x}_i)) - \gamma\sum_{i=1}^m p_i\phi\left(\frac{q_i}{p_i}\right)\right\} \\
&\leq \gamma^* r + \sup_{\substack{\mathbf{q}\geq\mathbf{0}\\ \mathbf{q}^T\mathbf{1}=1}} \sup_{\bar{\mathbf{q}}\in M_h(\mathbf{q})} -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a},\mathbf{x}_i)) - \gamma^*\sum_{i=1}^m p_i\phi\left(\frac{q_i}{p_i}\right).
\end{aligned}$$

Bringing $\gamma^* r$ to the left-hand side of the inequality, we obtain from (EC.24) that

$$\begin{aligned}
&\sup_{\substack{\mathbf{q}\geq\mathbf{0}\\ \mathbf{q}^T\mathbf{1}=1}} \sup_{\bar{\mathbf{q}}\in M_h(\mathbf{q})} -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}^*,\mathbf{x}_i)) - \gamma^*\sum_{i=1}^m p_i\phi\left(\frac{q_i}{p_i}\right) \\
&\leq \sup_{\substack{\mathbf{q}\geq\mathbf{0}\\ \mathbf{q}^T\mathbf{1}=1}} \sup_{\bar{\mathbf{q}}\in M_h(\mathbf{q})} -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a},\mathbf{x}_i)) - \gamma^*\sum_{i=1}^m p_i\phi\left(\frac{q_i}{p_i}\right),
\end{aligned}$$

for any $\mathbf{a}\in\mathcal{A}$. Hence, $\mathbf{a}^*$ is also a solution of (EC.23), where $\theta^* = \gamma^*$. $\square$

EC.5 A Visualization of the Various Shapes of the Uncertainty Set $\mathcal{U}_{\phi,h}(\mathbf{p})$



(a). $n = \infty$

(b). $n = 500$

(c). $n = 100$

(d). $n = 5$

Figure EC.1: Projections of the uncertainty set $\mathcal{U}_{\phi,h}(\mathbf{p})$ in (13) on the coordinates $(\bar{q}_1, \bar{q}_2)$, for $\mathbf{p} = (1/3, 1/3, 1/3)$ and $r = \frac{1}{n}\chi^2_{0.95,2}$, plotted for a range of values of the sample size n. We choose the modified chi-squared divergence function $\phi(t) = (t-1)^2$ and the second dual moment distortion function $h(p) = 1-(1-p)^2$. As n approaches 0, we observe that the uncertainty set grows and the projection eventually approaches the entire probability simplex in $\mathbb{R}^2$, the case in which the decision-maker is completely ambiguous w.r.t. $\mathbf{p}$.

EC.6 The Optimistic Dual Counterpart

Beck and Ben-Tal (2009) and Gorissen et al. (2014) have shown that a robust minimization problem with a compact convex uncertainty set can also be reformulated by considering the optimistic dual counterpart, obtained by maximizing the dual of its uncertain problem over all uncertain variables. In this section, we derive and reformulate the optimistic dual counterpart of the robust problem

$$\min_{\mathbf{a}\in\mathcal{A}, c\in\mathcal{C}} \left\{ \mathbf{a}^T\mathbf{d} + \zeta \cdot c \;\middle|\; \sup_{(\mathbf{q},\bar{\mathbf{q}})\in\mathcal{U}_{\phi,h}(\mathbf{p})} -\sum_{i=1}^{m} \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) \leq c \right\}, \tag{EC.25}$$

where

$$\mathcal{U}_{\phi,h}(\mathbf{p}) = \left\{ (\mathbf{q}, \bar{\mathbf{q}}) \in \mathbb{R}^{2m} \;\middle|\; \begin{array}{l} \sum_{i=1}^m q_i = \sum_{i=1}^m \bar{q}_i = 1 \\ \sum_{i=1}^m p_i \phi\left(\frac{q_i}{p_i}\right) \le r \\ \sum_{i\in J} \bar{q}_i \le h\left(\sum_{i\in J} q_i\right), \ \forall J \subset [m] \\ q_i, \bar{q}_i \ge 0, \ \forall i \in [m] \end{array} \right\}.$$

Here, we have either ($\mathcal{C} = \mathbb{R}, \mathbf{d} = \mathbf{0}, \zeta = 1$) or ($\mathcal{C} = \{c_0\}, \mathbf{d} \neq \mathbf{0}, \zeta = 0$), where $\mathbf{d} \in \mathbb{R}^{n_a}, c_0 \in \mathbb{R}$. Note that $\mathcal{U}_{\phi,h}(\mathbf{p})$ is a compact set since it is bounded and closed due to Assumption 3 that the functions $\phi, -h$ are lower-semicontinuous convex functions.

To derive the optimistic dual counterpart of (EC.25), we first consider its uncertain problem, which for a given $(\mathbf{q}, \bar{\mathbf{q}}) \in \mathcal{U}_{\phi,h}(\mathbf{p})$ is defined as

$$\begin{aligned} &\min_{\mathbf{a}\in\mathcal{A}, c\in\mathcal{C}} \left\{ \mathbf{a}^T\mathbf{d} + \zeta \cdot c \;\middle|\; -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) \le c \right\} \\ &= \min_{\mathbf{a}\in\mathbb{R}^I, c\in\mathcal{C}} \left\{ \mathbf{a}^T\mathbf{d} + \zeta \cdot c \;\middle|\; -\sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) \le c, f_i(\mathbf{a}) \le 0, i \in [L] \right\}. \end{aligned} \tag{EC.26}$$

Here, recall that $\mathcal{A}$ is a set represented by convex inequalities, hence we can express it as $\mathcal{A} = \{\mathbf{a} \in \mathbb{R}^I | f_i(\mathbf{a}) \le 0, i = 1, \ldots, L\}$, for some convex functions f_i's. Assume that (EC.26) satisfies Slater's condition and is bounded from below. Then, the dual of (EC.26) is given by

$$\max_{y_0,\ldots,y_L\ge 0} \inf_{\mathbf{a}\in\mathbb{R}^I, c\in\mathcal{C}} \left\{ \mathbf{a}^T\mathbf{d} + \zeta \cdot c - y_0 \sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) - y_0 c + \sum_{k=1}^L y_k f_k(\mathbf{a}) \right\}. \tag{EC.27}$$

The optimistic dual counterpart is defined by maximizing the dual over all uncertain variables in the uncertainty set, i.e.,

$$\sup_{(\mathbf{q},\bar{\mathbf{q}})\in\mathcal{U}_{\phi,h}(\mathbf{p})} \max_{y_0,\ldots,y_L\ge 0} \inf_{\mathbf{a}\in\mathbb{R}^I, c\in\mathcal{C}} \left\{ \mathbf{a}^T\mathbf{d} + \zeta \cdot c - y_0 \sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) - y_0 c + \sum_{k=1}^L y_k f_k(\mathbf{a}) \right\}. \tag{OD}$$

The following theorem states a reformulation of (OD).

Theorem EC.1. *The optimistic dual counterpart* (OD) *is equivalent to the following concave problem:*

$$\begin{aligned} \sup_{\substack{\mathbf{z},\bar{\mathbf{z}}\in\mathbb{R}^m_{\ge 0} \\ y_0,\ldots,y_L\ge 0 \\ \boldsymbol{\lambda}_1,\ldots,\boldsymbol{\lambda}_m,\boldsymbol{\eta}_1,\ldots,\boldsymbol{\eta}_L\in\mathbb{R}^I}} \quad & -\sum_{i=1}^m \bar{z}_i(-u\circ f)^*\left(\frac{\boldsymbol{\lambda}_i}{y_0\bar{q}_i}, \mathbf{x}_i\right) - \sum_{k=1}^L y_k (f_k)^*\left(\frac{\boldsymbol{\eta}_k}{y_k}\right) - y_0 c_0 \mathbb{1}_{\{\zeta=0\}} \\ \text{subject to } & \sum_{i=1}^m \boldsymbol{\lambda}_i = -\sum_{k=1}^L \boldsymbol{\eta}_k - \mathbf{d} \\ & \sum_{i=1}^m z_i = \sum_{i=1}^m \bar{z}_i = y_0 \\ & \sum_{i\in I} \bar{z}_i - y_0 h\left(\frac{\sum_{i\in I} z_i}{y_0}\right) \le 0, \forall I \in 2^N_-, \\ & \sum_{i=1}^m y_0 p_i \phi\left(\frac{z_i}{y_0 p_i}\right) - y_0 r \le 0, \end{aligned} \tag{EC.28}$$

where $\mathbf{z} = y_0\mathbf{q}$ *and* $\bar{\mathbf{z}} = y_0\bar{\mathbf{q}}$*. If Slater's condition holds for problem* (EC.28)*, then* (EC.25) *is equal to* (EC.28)*. Moreover, the KKT-vector of* (EC.28) *corresponding to the dual equality constraints* $\sum_{i=1}^m \boldsymbol{\lambda}_i = -\sum_{k=1}^L \boldsymbol{\eta}_k - \mathbf{d}$ *gives the optimal solution of* (EC.25)*.*

Proof. We have that

$$\begin{aligned}
&\inf_{\mathbf{a}\in\mathbb{R}^I, c\in\mathcal{C}} \left\{ \mathbf{a}^T\mathbf{d} + \zeta \cdot c - y_0 \sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) - y_0 c + \sum_{k=1}^L y_k f_k(\mathbf{a}) \right\} \\
&= \inf_{\mathbf{a}\in\mathbb{R}^I} \left\{ \mathbf{a}^T\mathbf{d} - y_0 \sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) + \sum_{k=1}^L y_k f_k(\mathbf{a}) \right\} + \inf_{c\in\mathcal{C}} (\zeta - y_0)c \\
&= -\sup_{\mathbf{a}\in\mathbb{R}^I} \left\{ -\mathbf{a}^T\mathbf{d} + y_0 \sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) - \sum_{k=1}^L y_k f_k(\mathbf{a}) \right\} + \inf_{c\in\mathcal{C}} (\zeta - y_0)c.
\end{aligned}$$

Note that $\mathcal{C}$ is either $\mathbb{R}$ (if $\zeta = 1$) or $\{c_0\}$ for some $c_0 \in \mathbb{R}$ (if $\zeta = 0$). Hence, we have that

$$\inf_{c\in\mathcal{C}} (\zeta - y_0)c = \begin{cases} -y_0 c_0 & \text{if } \mathcal{C} = c_0 \\ 0 & \text{if } \mathcal{C} = \mathbb{R},\ y_0 = 1 \\ -\infty & \text{else.} \end{cases}$$

We examine the supremum term. We have

$$\begin{aligned}
&\sup_{\mathbf{a}\in\mathbb{R}^I} \left\{ -\mathbf{a}^T\mathbf{d} + y_0 \sum_{i=1}^m \bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) - \sum_{k=1}^L y_k f_k(\mathbf{a}) \right\} \\
&= \sup_{\substack{\mathbf{a}, \mathbf{w}_1, \dots, \mathbf{w}_m, \\ \mathbf{v}_1, \dots, \mathbf{v}_L \in \mathbb{R}^I}} \left\{ -\mathbf{a}^T\mathbf{d} + y_0 \sum_{i=1}^m \bar{q}_i u(f(\mathbf{w}_i, \mathbf{x}_i)) - \sum_{k=1}^L y_k f_k(\mathbf{v}_k) \;\middle|\; \mathbf{w}_i = \mathbf{a},\ \mathbf{v}_k = \mathbf{a}, i \in [m], k \in [L] \right\} \\
&= \inf_{\substack{\boldsymbol{\lambda}_1, \dots \boldsymbol{\lambda}_m, \\ \boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_L \in \mathbb{R}^I}} \sup_{\substack{\mathbf{a}, \mathbf{w}_1, \dots, \mathbf{w}_m, \\ \mathbf{v}_1, \dots, \mathbf{v}_L \in \mathbb{R}^I}} \left\{ -\mathbf{a}^T\mathbf{d} + \sum_{i=1}^m \boldsymbol{\lambda}_i^T(\mathbf{w}_i - \mathbf{a}) + \sum_{k=1}^L \boldsymbol{\eta}_k^T(\mathbf{v}_k - \mathbf{a}) + y_0 \sum_{i=1}^m \bar{q}_i u(f(\mathbf{w}_i, \mathbf{x}_i)) \right. \\
&\qquad \left. - \sum_{k=1}^L y_k f_k(\mathbf{v}_k) \right\} \\
&= \inf_{\substack{\boldsymbol{\lambda}_1, \dots \boldsymbol{\lambda}_m, \\ \boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_L \in \mathbb{R}^I}} \sup_{\substack{\mathbf{a}, \mathbf{w}_1, \dots, \mathbf{w}_m, \\ \mathbf{v}_1, \dots, \mathbf{v}_L \in \mathbb{R}^I}} \left\{ -\left(\sum_{i=1}^m \boldsymbol{\lambda}_i + \sum_{k=1}^L \boldsymbol{\eta}_k + \mathbf{d} \right)^T \mathbf{a} + \sum_{i=1}^m \boldsymbol{\lambda}_i^T \mathbf{w}_i - y_0 \bar{q}_i (-u)(f(\mathbf{w}_i, \mathbf{x}_i)) \right. \\
&\qquad \left. + \sum_{k=1}^L \boldsymbol{\eta}_k^T \mathbf{v}_k - y_k f_k(\mathbf{v}_k) \right\} \\
&= \inf_{\substack{\boldsymbol{\lambda}_1, \dots \boldsymbol{\lambda}_m, \\ \boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_L \in \mathbb{R}^I}} \left\{ \sum_{i=1}^m y_0 \bar{q}_i (-u \circ f)^* \left(\frac{\boldsymbol{\lambda}_i}{y_0 \bar{q}_i}, \mathbf{x}_i \right) + \sum_{k=1}^L y_k (f_k)^* \left(\frac{\boldsymbol{\eta}_k}{y_k} \right) \;\middle|\; \sum_{i=1}^m \boldsymbol{\lambda}_i = -\sum_{k=1}^L \boldsymbol{\eta}_k - \mathbf{d} \right\},
\end{aligned}$$

where $(-u \circ f)$ denotes the composition of the two functions. A change of variables $\mathbf{z} = y_0\mathbf{q}$ and $\bar{\mathbf{z}} = y_0\bar{\mathbf{q}}$ and a multiplication of the constraints in $\mathcal{U}_{\phi,h}(\mathbf{p})$ by y_0 yields the desired statement. □

EC.7 Optimization of Rank-Dependent Models in the Constraint

We describe in detail the adaptation of Algorithm 1 to (P-constraint) in Algorithm EC.1.

Algorithm EC.1 Cutting-Plane Method with RDU Constraint

1: Start with $\mathcal{U}_1 = \{(\mathbf{p}, \mathbf{p})\}$. Fix a tolerance parameter $\epsilon_{\text{tol}} > 0$.

2: At the j-th iteration, solve the following problem with the uncertainty set $\mathcal{U}_j$:

$$\min_{\mathbf{a}\in\mathcal{A}} \left\{ g(\mathbf{a}) \;\middle|\; \sup_{(\mathbf{q},\bar{\mathbf{q}})\in\mathcal{U}_j} \sum_{i=1}^{m} -\bar{q}_i u(f(\mathbf{a}, \mathbf{x}_i)) \leq c \right\}. \tag{EC.29}$$

3: Let $\mathbf{a}_j$ be the optimal solution of (20). Determine the ranking:

$$-u(f(\mathbf{a}_j, \mathbf{x}_{(1)})) \leq \ldots \leq -u(f(\mathbf{a}_j, \mathbf{x}_{(m)})).$$

Then, solve the optimization problem (19), which gives an optimal objective value v_j and a solution $(\mathbf{q}_j^*, \bar{\mathbf{q}}_j^*)$.

4: If $v_j - c \leq \epsilon_{\text{tol}}$, then the solution is accepted and the process is terminated.

5: If not, set $\mathcal{U}_{j+1} = \mathcal{U}_j \cup \{(\mathbf{q}_j^*, \bar{\mathbf{q}}_j^*)\}$ and repeat steps 2–5.

EC.8 SOS2-Constraints Formulation

Let $u : [l_0, u_0] \to \mathbb{R}$ be a non-decreasing, piecewise-linear utility function defined on some interval $[l_0, u_0]$ that contains the image set $\{\mathbf{a}^T\mathbf{x}_i \mid \mathbf{a} \in \mathcal{A}, i \in [m]\}$. Let $\{t_j, u(t_j)\}_{j=1}^K$ be the support points of u, where $l_0 = t_1 < \cdots < t_K = u_0$. Then, the constraints in (34) can be formulated using the following bilinear and SOS2 constraints:

$$\begin{cases} \beta \cdot h(p^0) + \sum_{k=1}^{K_1} \nu_k b_k^{(1)} + \sum_{i=1}^m \sum_{k=1}^{K_1} \lambda_{ik} l_k^{(1)} p_i - \sum_{i=1}^m \bar{q}_i z_i \leq c \\ -z_i - \beta - \sum_{k=1}^{K_1} \lambda_{ik} \leq 0,\ \forall i \in [m] \\ \lambda_{ik} \leq \nu_k,\ \forall i \in [m],\ \forall k \in [K_1] \\ \bar{q}_i \leq l_k^{(2)} p_i + t_{ik},\ \forall i \in [m],\ \forall k \in [K_2] \\ \sum_{i=1}^m t_{ik} \leq b_k^{(2)},\ \forall k \in [K_2] \\ \sum_{i=1}^m \bar{q}_i = \bar{h}(1 - p^0) \\ \mathbf{a}^T\mathbf{x}_i = \sum_{j=1}^K \tilde{\lambda}_{ij} t_j,\ \forall i \in [m] \\ z_i = \sum_{j=1}^K \tilde{\lambda}_{ij} u(t_j),\ \forall i \in [m] \\ \sum_{j=1}^K \tilde{\lambda}_{ij} = 1,\ \forall i \in [m] \\ \tilde{\lambda}_{ij} \geq 0,\ \text{SOS2},\ \forall i \in [m],\ \forall j \in [K]. \end{cases} \tag{EC.30}$$

EC.9 The Hit-and-Run Algorithm

In this appendix, we explain how the Hit-and-Run algorithm is applied to generate a uniform sample on the ϕ-divergence set. Let S be an open subset in $\mathbb{R}^d$. The general procedure of the Hit-and-Run algorithm is fairly simple. Start with an interior point $x_0 \in S$. Choose uniformly a random direction $u \in \partial\mathcal{D}$ where $\partial\mathcal{D} \triangleq \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$. Draw uniformly a scalar $\lambda \in \{\lambda \in \mathbb{R} : x_0 + \lambda u \in S\}$ and then update x_0 with $x_0 + \lambda u$. It is shown by Bélisle et al. (1993) that for a convex set S, this sampling process will converge to a uniform sample on S.

To apply the Hit-and-Run algorithm to the ϕ-divergence uncertainty set, we first re-parametrize the set. Recall the definition of the ϕ-divergence set (8). Using the equality $\mathbf{q}^T\mathbf{1} = 1$, it is clear that we can parametrize the ϕ-divergence set with the following set:

$$\tilde{\mathcal{D}}_\phi(\mathbf{q}|\mathbf{p}, r) \triangleq \{\mathbf{q} \in \mathbb{R}^{m-1} | \mathbf{q} \geq \mathbf{0}, 1 - \sum_{i=1}^{m-1} q_i \geq 0, \sum_{i=1}^{m-1} p_i \phi\left(\frac{q_i}{p_i}\right) + p_m \phi\left(\frac{1 - \sum_{i=1}^{m-1} q_i}{p_m}\right) \leq r\}. \tag{EC.31}$$

Since this is a set with convex inequalities, and we assume Slater's condition, its interior is an open convex set. Hence, we can apply the Hit-and-Run algorithm to obtain a uniform sample of the interior of $\tilde{\mathcal{D}}_\phi(\mathbf{q}|\mathbf{p}, r)$ and thus also the interior of $\mathcal{D}_\phi(\mathbf{p}, r)$ due to their one-to-one correspondence.

The precise procedure is the following.

- We start with an interior point $\mathbf{q}_k \in \text{int}\tilde{\mathcal{D}}_\phi(\mathbf{q}|\mathbf{p}, r)$, where we set $\mathbf{q}_0 = (p_1, \ldots, p_{m-1})$. Then, we draw a uniform element $\mathbf{u} \in \partial\mathcal{D}$ of dimension $d = m - 1$. This can be done by drawing d standard normal i.i.d. samples $\mathbf{X} = (X_1, \ldots, X_d)$ and setting $\mathbf{u} = \frac{\mathbf{X}}{\|\mathbf{X}\|_2}$, where $\|.\|_2$ is the Euclidean 2-norm. Then, $\mathbf{u}$ follows a uniform distribution on $\partial\mathcal{D}$.

- Given $\mathbf{u}$ and $\mathbf{q}_k$, we determine the set $\Lambda \triangleq \{\lambda \in \mathbb{R} : \mathbf{q}_k + \lambda\mathbf{u} \in \tilde{\mathcal{D}}_\phi(\mathbf{q}|\mathbf{p}, r)\}$. Note that since $\tilde{\mathcal{D}}_\phi(\mathbf{q}|\mathbf{p}, r)$ is convex, bounded and closed, there exist $\lambda_{\min}, \lambda_{\max}$ such that $[\lambda_{\min}, \lambda_{\max}] = \Lambda$. We may find both values using a bisection search. Indeed, since $\mathbf{q}_k \in \text{int}\tilde{\mathcal{D}}_\phi(\mathbf{q}|\mathbf{p}, r)$, there exists a sufficiently small $\lambda_0 > 0$ such that $\lambda_0 \in \Lambda$. Take a sufficiently large $\lambda_1 > \lambda_0$ such that $\lambda_1 \notin \Lambda$, which also exists since Λ is bounded. Then, perform the bisection search on $[\lambda_0, \lambda_1]$ to find $\lambda_{\max}$. Do the same for $\lambda_{\min}$.

- We draw a random $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ and update $\mathbf{q}_{k+1} = \mathbf{q}_k + \lambda\mathbf{u}$. We then repeat the process.

References

Beck, A. and Ben-Tal, A. (2009). Duality in robust optimization: Primal worst equals dual best. *Operations Research Letters*, 37:1–6.

Ben-Tal, A., den Hertog, D., de Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357.

Ben-Tal, A. and Nemirovski, A. (2019). *Lectures on Modern Convex Optimization.* SIAM.

Berge, C. (1963). *Topological Spaces: Including a Treatment of Multi-Valued Functions, Vector Spaces, and Convexity.* Courier Corporation.

Bélisle, C. J., Romeijn, H. E., and Smith, R. L. (1993). Hit-and-Run algorithms for generating multivariate distributions. *Mathematics of Operations Research*, 18(2):255–266.

Chares, P. R. (2009). *Cones and Interior-Point Algorithms for Structured Convex Optimization Involving Powers and Exponentials.* Ecole polytechnique de Louvain, Université Catholique de Louvain.

Cherny, A. and Madan, D. (2009). New measures for performance evaluation. *The Review of Financial Studies*, 22(7):2571–2606.

Denneberg, D. (1990a). Distorted probabilities and insurance premiums. *Methods of Operations Research*, 63:3–5.

Denneberg, D. (1990b). Premium calculation: Why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin: The Journal of the IAA*, 20(2):181–190.

Denneberg, D. (1994). *Non-Additive Measure and Integral*. Springer.

Eeckhoudt, L. R., Laeven, R. J., and Schlesinger, H. (2020). Risk apportionment: The dual story. *Journal of Economic Theory*, 185:104971.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance*. Berlin: De Gruyter, 4th edition.

Gorissen, B. L., Blanc, H., Ben-Tal, A., and den Hertog, D. (2014). Technical Note–Deriving robust and globalized robust solutions of uncertain linear programs with general convex uncertainty sets. *Operations Research*, 62(3):672–679.

Muliere, P. and Scarsini, M. (1989). A note on stochastic dominance and inequality measures. *Journal of Economic Theory*, 49(2):314–323.

Mutapcic, A. and Boyd, S. (2009). Cutting-set methods for robust convex optimization with pessimizing oracles. *Optimization Methods & Software*, 24(3):381–406.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, NJ.

Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17(1):43–54.