Robust Risk-Aware Reinforcement Learning

Sebastian Jaimungal   Silvana Pesenti   Ye Sheng Wang   Hariom Tatsat

sebastian.statistics.utoronto.ca

May 30, 2022

Department of Statistical Sciences
University of Toronto

Reinforcement Learning

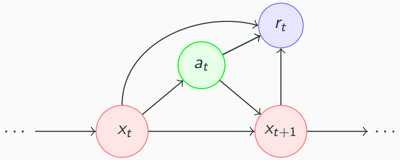


Figure 1: Directed graph representation of the stochastic control problem.

- Total reward $Z^\theta = \sum_{t=0}^{T-1} \gamma^t \, r_t$
- θ parameterise the policy: $a_t = \pi_\theta(t, x_t)$ or $a_t \overset{\mathbb{P}}{\sim} \pi_\theta(t, x_t)$

Standard RL: *risk-neutral objective* function of a cost

$$\min_\theta \mathbb{E}\left[Z^\theta\right].$$

Risk-aware RL: *risk measure ρ* of the cost Z

$$\min_\theta \rho(Z^\theta) \quad \text{or} \quad \min_\theta \mathbb{E}\left[Z^\theta\right] \text{ subj. to } \rho(Z^\theta) \leq R$$

Reinforcement Learning

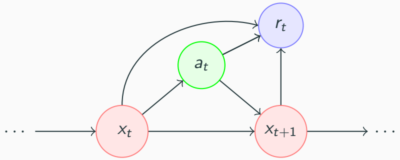


Figure 1: Directed graph representation of the stochastic control problem.

- Total reward $Z^\theta = \sum_{t=0}^{T-1} \gamma^t \, r_t$
- θ parameterise the policy: $a_t = \pi_\theta(t, x_t)$ or $a_t \overset{\mathbb{P}}{\sim} \pi_\theta(t, x_t)$

Standard RL: *risk-neutral objective* function of a cost

$$\min_\theta \mathbb{E}\left[Z^\theta\right].$$

Risk-aware RL: *risk measure ρ* of the cost Z

$$\min_\theta \rho(Z^\theta) \quad \text{or} \quad \min_\theta \mathbb{E}\left[Z^\theta\right] \text{ subj. to } \rho(Z^\theta) \leq R$$

Reinforcement Learning

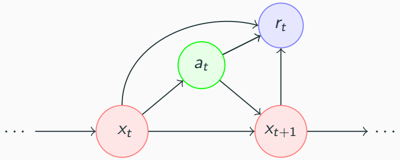


Figure 1: Directed graph representation of the stochastic control problem.

- Total reward $Z^\theta = \sum_{t=0}^{T-1} \gamma^t \, r_t$
- θ parameterise the policy: $a_t = \pi_\theta(t, x_t)$ or $a_t \overset{\mathbb{P}}{\sim} \pi_\theta(t, x_t)$

Standard RL: *risk-neutral objective* function of a cost

$$\min_\theta \mathbb{E}\left[Z^\theta\right].$$

Risk-aware RL: *risk measure* ρ of the cost Z

$$\min_\theta \rho(Z^\theta) \quad \text{or} \quad \min_\theta \mathbb{E}\left[Z^\theta\right] \text{ subj. to } \rho(Z^\theta) \leq R$$

Motivation

- Agents often weigh outcomes unequally
- Probabilities of outcomes are often distorted
- *Rank dependent expected utility* (RDEU) / Yaari's dual Theory

$$\mathcal{R}_{g}^{U}[Y] := \int_{-\infty}^{0} \Big\{ 1 - g(\mathbb{P}(U(Y) > y)) \Big\} \, dy - \int_{0}^{+\infty} g(\mathbb{P}(U(Y) > y)) \, dy$$

  - g is an increasing probability distortion
  - U is a concave utility

- Helps explain the Allais paradox...
  - A: 61% chance to win \$1.2M OR 63% chance to win \$1M
  - B: 98% chance to win \$1.2M OR 100% chance to win \$1M
- $U(x) = x$, recovers distortion risk-measures
- $g(s) = s$, recovers expected utility
- g may be inverse-S shaped

Motivation

- Agents often weigh outcomes unequally
- Probabilities of outcomes are often distorted
- *Rank dependent expected utility* (RDEU) / Yaari's dual Theory

$$\mathcal{R}_{g}^{U}[Y] := \int_{-\infty}^{0} \left\{ 1 - g(\mathbb{P}(U(Y) > y)) \right\} dy - \int_{0}^{+\infty} g(\mathbb{P}(U(Y) > y))\, dy$$

  - g is an increasing probability distortion
  - U is a concave utility

- Helps explain the Allais paradox...
  - A: 61% chance to win \$1.2M OR 63% chance to win \$1M
  - B: 98% chance to win \$1.2M OR 100% chance to win \$1M
- $U(x) = x$, recovers distortion risk-measures
- $g(s) = s$, recovers expected utility
- g may be inverse-S shaped

Motivation

- Agents often weigh outcomes unequally
- Probabilities of outcomes are often distorted
- *Rank dependent expected utility* (RDEU) / Yaari's dual Theory

$$\mathcal{R}_{\boldsymbol{g}}^{\boldsymbol{U}}[Y] := \int_{-\infty}^{0} \Big\{ 1 - \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y)) \Big\}\, dy - \int_{0}^{+\infty} \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y))\, dy$$

  - $\boldsymbol{g}$ is an increasing probability distortion
  - $\boldsymbol{U}$ is a concave utility
- Helps explain the Allais paradox...
  - A: 61% chance to win \$1.2M OR 63% chance to win \$1M
  - B: 98% chance to win \$1.2M OR 100% chance to win \$1M
- $U(x) = x$, recovers distortion risk-measures
- $g(s) = s$, recovers expected utility
- g may be inverse-S shaped

Motivation

- Agents often weigh outcomes unequally
- Probabilities of outcomes are often distorted
- *Rank dependent expected utility* (RDEU) / Yaari's dual Theory

$$\mathcal{R}_{\boldsymbol{g}}^{\boldsymbol{U}}[Y] := \int_{-\infty}^{0} \left\{ 1 - \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y)) \right\} dy - \int_{0}^{+\infty} \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y))\, dy$$

  - $\boldsymbol{g}$ is an increasing probability distortion
  - $\boldsymbol{U}$ is a concave utility

- Helps explain the Allais paradox...
  - A: 61% chance to win \$1.2M OR 63% chance to win \$1M
  - B: 98% chance to win \$1.2M OR 100% chance to win \$1M
- $U(x) = x$, recovers distortion risk-measures
- $g(s) = s$, recovers expected utility
- g may be inverse-S shaped

Motivation

- Agents often weigh outcomes unequally
- Probabilities of outcomes are often distorted
- *Rank dependent expected utility* (RDEU) / Yaari's dual Theory

$$\mathcal{R}_{\boldsymbol{g}}^{\boldsymbol{U}}[Y] := \int_{-\infty}^{0} \left\{ 1 - \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y)) \right\} dy - \int_{0}^{+\infty} \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y))\, dy$$

  - $\boldsymbol{g}$ is an increasing probability distortion
  - $\boldsymbol{U}$ is a concave utility

- Helps explain the Allais paradox...
  - A: 61% chance to win \$1.2M OR 63% chance to win \$1M
  - B: 98% chance to win \$1.2M OR 100% chance to win \$1M
- $U(x) = x$, recovers distortion risk-measures
- $g(s) = s$, recovers expected utility
- g may be inverse-S shaped

Motivation

- Agents often weigh outcomes unequally
- Probabilities of outcomes are often distorted
- *Rank dependent expected utility* (RDEU) / Yaari's dual Theory

$$\mathcal{R}_{\boldsymbol{g}}^{\boldsymbol{U}}[Y] := \int_{-\infty}^{0} \Big\{ 1 - \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y)) \Big\}\, dy - \int_{0}^{+\infty} \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y))\, dy$$

  - $\boldsymbol{g}$ is an increasing probability distortion
  - $\boldsymbol{U}$ is a concave utility

- Helps explain the Allais paradox...
  - A: 61% chance to win \$1.2M OR 63% chance to win \$1M
  - B: 98% chance to win \$1.2M OR 100% chance to win \$1M
- $U(x) = x$, recovers distortion risk-measures
- $g(s) = s$, recovers expected utility
- g may be inverse-S shaped

Motivation

- Agents often weigh outcomes unequally
- Probabilities of outcomes are often distorted
- *Rank dependent expected utility* (RDEU) / Yaari's dual Theory

$$\mathcal{R}_{\boldsymbol{g}}^{\boldsymbol{U}}[Y] := \int_{-\infty}^{0} \left\{ 1 - \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y)) \right\} dy - \int_{0}^{+\infty} \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y)) \, dy$$

  - $\boldsymbol{g}$ is an increasing probability distortion
  - $\boldsymbol{U}$ is a concave utility

- Helps explain the Allais paradox...
  - A: 61% chance to win \$1.2M OR 63% chance to win \$1M
  - B: 98% chance to win \$1.2M OR 100% chance to win \$1M
- $U(x) = x$, recovers distortion risk-measures
- $g(s) = s$, recovers expected utility
- g may be inverse-S shaped

Motivation

- Agents often weigh outcomes unequally
- Probabilities of outcomes are often distorted
- *Rank dependent expected utility* (RDEU) / Yaari's dual Theory

$$\mathcal{R}_{\boldsymbol{g}}^{\boldsymbol{U}}[Y] := \int_{-\infty}^{0} \Big\{ 1 - \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y)) \Big\} \, dy - \int_{0}^{+\infty} \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y)) \, dy$$

  - $\boldsymbol{g}$ is an increasing probability distortion
  - $\boldsymbol{U}$ is a concave utility

- Helps explain the Allais paradox...
  - A: 61% chance to win \$1.2M OR 63% chance to win \$1M
  - B: 98% chance to win \$1.2M OR 100% chance to win \$1M
- $U(x) = x$, recovers distortion risk-measures
- $g(s) = s$, recovers expected utility
- g may be inverse-S shaped

Motivation

- Agents often weigh outcomes unequally
- Probabilities of outcomes are often distorted
- *Rank dependent expected utility* (RDEU) / Yaari's dual Theory

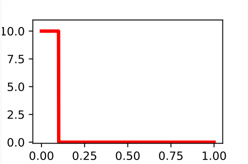
$$\mathcal{R}_{\boldsymbol{g}}^{\boldsymbol{U}}[Y] := \int_{-\infty}^{0} \left\{ 1 - \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y)) \right\} dy - \int_{0}^{+\infty} \boldsymbol{g}(\mathbb{P}(\boldsymbol{U}(Y) > y))\, dy$$

  - $\boldsymbol{g}$ is an increasing probability distortion
  - $\boldsymbol{U}$ is a concave utility

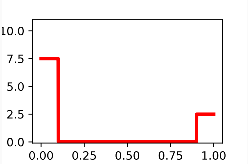
- Helps explain the Allais paradox...
  - A: 61% chance to win \$1.2M OR 63% chance to win \$1M
  - B: 98% chance to win \$1.2M OR 100% chance to win \$1M
- $U(x) = x$, recovers distortion risk-measures
- $g(s) = s$, recovers expected utility
- g may be inverse-S shaped

Example: $\alpha - \beta$ risk measure

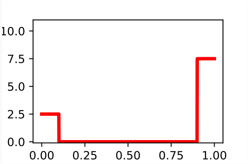
- α-β risk measure $\gamma(u) = \frac{1}{\eta}\left(p\,\mathbb{1}_{\{u\leq\alpha\}} + (1-p)\,\mathbb{1}_{\{u>\beta\}}\right)$



$p = 1$



$p > \frac{1}{2}$



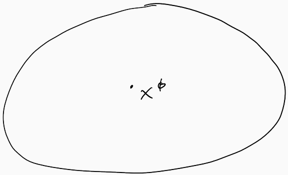
$p < \frac{1}{2}$

- α-β risk measure is U-shaped and contains several notable special cases
  - $p = 1$ corresponds to the TVaR at level α
  - $p > \frac{1}{2}$ emphasises losses relative to gains
  - $p < \frac{1}{2}$ emphasises gains relative to losses

non-Robust Problem Setup

The risk-aware RL problems we address are

$$\inf_{\phi \in \varphi} \mathcal{R}_g^U[X^\phi] \qquad \text{(P)}$$



For example:

- X^ϕ is the terminal wealth of a self-financing trading strategy with trading frictions

$$X^\phi = \int_0^T a(t, S_t, \phi)\, dS_t - c \int_0^T |a(t, S_t, \phi)|\, dt$$

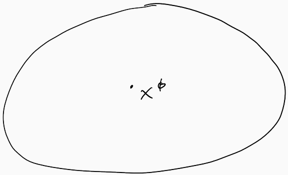
with

$$dS_t = \kappa(\theta - S_t) + \sigma\, dW_t$$

non-Robust Problem Setup

The risk-aware RL problems we address are

$$\inf_{\phi \in \varphi} \mathcal{R}_g^U[X^\phi] \qquad \text{(P)}$$



For example:

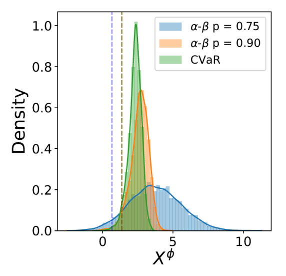
- X^ϕ is the terminal wealth of a self-financing trading strategy with trading frictions

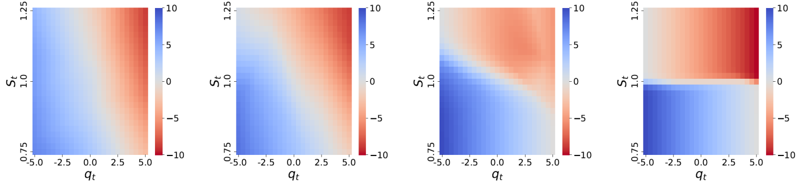
$$X^\phi = \int_0^T a(t, S_t, \phi)\, dS_t - c \int_0^T |a(t, S_t, \phi)|\, dt$$

with

$$dS_t = \kappa(\theta - S_t) + \sigma\, dW_t$$

Example: Statistical Arbitrage

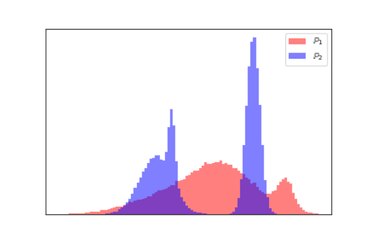


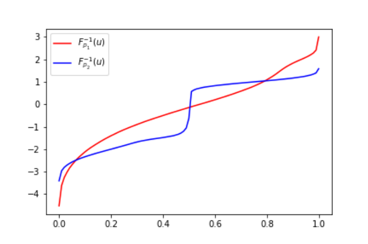


$\mathfrak{p} = 1$ $\mathfrak{p} = 0.9$ $\mathfrak{p} = 0.75$ risk-neutral

Robustification

- Models are often approximations of true dynamics... thus, we aim to *robustify decisions* by seeking over a *Wasserstein Ball*





$$d_p[X\,,\,Y] := \inf_{\chi\in\Pi(F_X,F_Y)} \left(\int_{\mathbb{R}^2} |x-y|^p\,\chi(dx,dy)\right)^{\frac{1}{p}}$$

$$= \left(\int_0^1 \left(F_X^{-1}(u) - F_Y^{-1}(u)\right)^p\,du\right)^{\frac{1}{p}}$$

Problem Setup

The class of robust risk-aware RL problems we address are

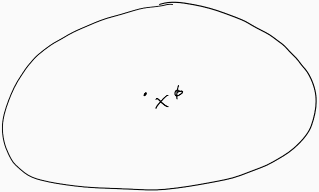
$$\inf_{\phi \in \varphi} \sup_{\theta \in \vartheta_\phi} \mathcal{R}_g^U[X^\theta], \qquad \text{where} \qquad \vartheta_\phi := \left\{ \theta \in \vartheta : d_p[X^\theta, X^\phi] \le \varepsilon \right\} \quad \text{(P)}$$

- The outer problem aims to optimise over strategies parametrised by ϕ
- The inner problem aims to robustify over alternates parametrised by θ
- We assume $X^\theta = H_\theta(X^\phi, Y)$, Y is other sources of randomness

Problem Setup

The class of robust risk-aware RL problems we address are

$$\inf_{\phi \in \varphi} \sup_{\theta \in \vartheta_\phi} \mathcal{R}_g^U[X^\theta], \qquad \text{where} \qquad \vartheta_\phi := \left\{ \theta \in \vartheta : d_p[X^\theta, X^\phi] \le \varepsilon \right\} \quad \text{(P)}$$

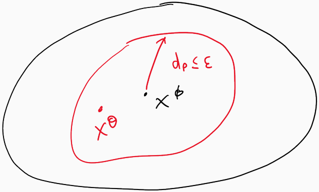


- The outer problem aims to optimise over strategies parametrised by ϕ
- The inner problem aims to robustify over alternates parametrised by θ
- We assume $X^\theta = H_\theta(X^\phi, Y)$, Y is other sources of randomness

Problem Setup

The class of robust risk-aware RL problems we address are

$$\inf_{\phi \in \varphi} \sup_{\theta \in \vartheta_\phi} \mathcal{R}_g^U[X^\theta]\,, \qquad \text{where} \qquad \vartheta_\phi := \left\{\theta \in \vartheta : d_p[X^\theta, X^\phi] \le \varepsilon\right\} \quad \text{(P)}$$

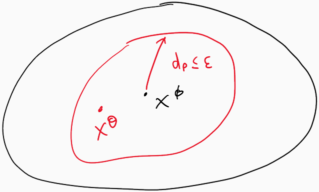


- The outer problem aims to optimise over strategies parametrised by ϕ
- The inner problem aims to robustify over alternates parametrised by θ
- We assume $X^\theta = H_\theta(X^\phi, Y)$, Y is other sources of randomness

Problem Setup

The class of robust risk-aware RL problems we address are

$$\inf_{\phi\in\varphi}\ \sup_{\theta\in\vartheta_\phi} \mathcal{R}_g^U[X^\theta]\,, \qquad \text{where} \qquad \vartheta_\phi := \left\{\theta \in \vartheta : d_p[X^\theta, X^\phi] \leq \varepsilon\right\} \quad \text{(P)}$$



- The outer problem aims to optimise over strategies parametrised by ϕ
- The inner problem aims to robustify over alternates parametrised by θ
- We assume $X^\theta = H_\theta(X^\phi, Y)$, Y is other sources of randomness

Problem Setup: Examples

- Robust Risk-Aware Portfolio Allocation [PW07; EK18]
  - φ a probability simplex – weights in traded assets (no short-selling)
  - $X^\phi = \phi^\intercal R$ total return
- Portfolio Optimisation within a Wasserstein Ball [PJ20]
  - φ is a singleton representing a reference portfolio
  - budget constraint on alternatives
  - Utility is linear, g arbitrary
- Robust Risk-Aware Statistical Arbitrage [CDJ17; LS21]
  - $\varphi = [-a, a]^{NT}$ – buy/sell actions of each asset
  - $X^\phi = \sum_{i=1}^{T} q_i^{\phi\intercal}(X_i - X_{i-1})$ total Profit & Loss
- Robust Barrier Option Hedging [CO11]
- Adversarial attacks

Problem Setup: Examples

- Robust Risk-Aware Portfolio Allocation [PW07; EK18]
  - φ a probability simplex – weights in traded assets (no short-selling)
  - $X^{\phi} = \phi^{\intercal} R$ total return
- Portfolio Optimisation within a Wasserstein Ball [PJ20]
  - φ is a singleton representing a reference portfolio
  - budget constraint on alternatives
  - Utility is linear, g arbitrary
- Robust Risk-Aware Statistical Arbitrage [CDJ17; LS21]
  - $\varphi = [-a, a]^{NT}$ – buy/sell actions of each asset
  - $X^{\phi} = \sum_{i=1}^{T} q_i^{\phi\intercal} (X_i - X_{i-1})$ total Profit & Loss
- Robust Barrier Option Hedging [CO11]
- Adversarial attacks

Problem Setup: Examples

- Robust Risk-Aware Portfolio Allocation [PW07; EK18]
  - φ a probability simplex – weights in traded assets (no short-selling)
  - $X^{\phi} = \phi^{\intercal} R$ total return
- Portfolio Optimisation within a Wasserstein Ball [PJ20]
  - φ is a singleton representing a reference portfolio
  - budget constraint on alternatives
  - Utility is linear, g arbitrary
- Robust Risk-Aware Statistical Arbitrage [CDJ17; LS21]
  - $\varphi = [-a, a]^{NT}$ – buy/sell actions of each asset
  - $X^{\phi} = \sum_{i=1}^{T} q_i^{\phi \intercal} (X_i - X_{i-1})$ total Profit & Loss
- Robust Barrier Option Hedging [CO11]
- Adversarial attacks

Problem Setup: Examples

- Robust Risk-Aware Portfolio Allocation [PW07; EK18]
  - φ a probability simplex – weights in traded assets (no short-selling)
  - $X^{\phi} = \phi^{\intercal} R$ total return
- Portfolio Optimisation within a Wasserstein Ball [PJ20]
  - φ is a singleton representing a reference portfolio
  - budget constraint on alternatives
  - Utility is linear, g arbitrary
- Robust Risk-Aware Statistical Arbitrage [CDJ17; LS21]
  - $\varphi = [-a, a]^{NT}$ – buy/sell actions of each asset
  - $X^{\phi} = \sum_{i=1}^{T} q_i^{\phi\intercal} (X_i - X_{i-1})$ total Profit & Loss
- Robust Barrier Option Hedging [CO11]
- Adversarial attacks

Problem Setup: Examples

- Robust Risk-Aware Portfolio Allocation [PW07; EK18]
  - φ a probability simplex – weights in traded assets (no short-selling)
  - $X^\phi = \phi^\intercal R$ total return
- Portfolio Optimisation within a Wasserstein Ball [PJ20]
  - φ is a singleton representing a reference portfolio
  - budget constraint on alternatives
  - Utility is linear, g arbitrary
- Robust Risk-Aware Statistical Arbitrage [CDJ17; LS21]
  - $\varphi = [-a, a]^{NT}$ – buy/sell actions of each asset
  - $X^\phi = \sum_{i=1}^{T} q_i^{\phi\intercal} (X_i - X_{i-1})$ total Profit & Loss
- Robust Barrier Option Hedging [CO11]
- Adversarial attacks

Augment Lagrangian

- We employ the *Augmented Lagrangian* approach [BM14] for optimisation

$$L[\theta, \phi] = \mathcal{R}_g^U[X^\theta] + \lambda\, c[X^\theta, X^\phi] + \tfrac{\mu}{2} \left(c[X^\theta, X^\phi]\right)^2,$$

  - $c[X^\theta, X^\phi] := \left((d_p[X^\theta, X^\phi])^p - \varepsilon^p\right)_+$ is the p-Wasserstein distance error
- Update rules
  - $\lambda \leftarrow \lambda + \mu\, c[X^\theta, X^\phi]$
  - $\mu \leftarrow a\, \mu$, and $a > 1$

Augment Lagrangian

- We employ the *Augmented Lagrangian* approach [BM14] for optimisation

$$L[\theta, \phi] = \mathcal{R}_g^U[X^\theta] + \lambda\, c[X^\theta, X^\phi] + \tfrac{\mu}{2}\left(c[X^\theta, X^\phi]\right)^2,$$

  - $c[X^\theta, X^\phi] := \left((d_p[X^\theta, X^\phi])^p - \varepsilon^p\right)_+$ is the p-Wasserstein distance error
- Update rules
  - $\lambda \leftarrow \lambda + \mu\, c[X^\theta, X^\phi]$
  - $\mu \leftarrow a\,\mu$, and $a > 1$

Augment Lagrangian

- We employ the *Augmented Lagrangian* approach [BM14] for optimisation

$$L[\theta, \phi] = \mathcal{R}_g^U[X^\theta] + \lambda\, c[X^\theta, X^\phi] + \tfrac{\mu}{2}\left(c[X^\theta, X^\phi]\right)^2,$$

  - $c[X^\theta, X^\phi] := \left((d_p[X^\theta, X^\phi])^p - \varepsilon^p\right)_+$ is the p-Wasserstein distance error
- Update rules
  - $\lambda \leftarrow \lambda + \mu\, c[X^\theta, X^\phi]$
  - $\mu \leftarrow a\, \mu$, and $a > 1$

Gradients

- Policy gradient [SMSM00] aims to optimise by updating parameters by $\theta \leftarrow \theta + \eta \nabla_\theta \rho(Z^\theta)$
- Risk-measures in the literature: (randomised policies)
  - Expected utility

$$\nabla_\theta \rho(Z^\theta) = \mathbb{E}^{\mathbb{P}} \left[\nabla_\theta \log \pi_\theta(a|x) \big|_{a=a^\theta} U(Z^\theta) \right]$$

  - Coherent risk measures [TCGM15]

$$\nabla_\theta \rho(Z^\theta) = \mathbb{E}^{\mathbb{P}^*} \left[\nabla_\theta \log \pi_\theta(a|x) \big|_{a=a^\theta} (Z^\theta - \lambda^*) \right]$$

Gradients

- Policy gradient [SMSM00] aims to optimise by updating parameters by $\theta \leftarrow \theta + \eta \nabla_\theta \rho(Z^\theta)$
- Risk-measures in the literature: (randomised policies)
  - Expected utility

$$\nabla_\theta \rho(Z^\theta) = \mathbb{E}^{\mathbb{P}} \left[\nabla_\theta \log \pi_\theta(a|x) \Big|_{a=a^\theta} U(Z^\theta) \right]$$

  - Coherent risk measures [TCGM15]

$$\nabla_\theta \rho(Z^\theta) = \mathbb{E}^{\mathbb{P}^*} \left[\nabla_\theta \log \pi_\theta(a|x) \Big|_{a=a^\theta} (Z^\theta - \lambda^*) \right]$$

Gradients

- Policy gradient [SMSM00] aims to optimise by updating parameters by $\theta \leftarrow \theta + \eta \nabla_\theta \rho(Z^\theta)$
- Risk-measures in the literature: (randomised policies)
  - Expected utility

$$\nabla_\theta \rho(Z^\theta) = \mathbb{E}^{\mathbb{P}} \left[\nabla_\theta \log \pi_\theta(a|x) \Big|_{a=a^\theta} U(Z^\theta) \right]$$

  - Coherent risk measures [TCGM15]

$$\nabla_\theta \rho(Z^\theta) = \mathbb{E}^{\mathbb{P}^*} \left[\nabla_\theta \log \pi_\theta(a|x) \Big|_{a=a^\theta} (Z^\theta - \lambda^*) \right]$$

Gradients

- Policy gradient [SMSM00] aims to optimise by updating parameters by $\theta \leftarrow \theta + \eta \nabla_\theta \rho(Z^\theta)$
- Risk-measures in the literature: (randomised policies)
  - Expected utility

$$\nabla_\theta \rho(Z^\theta) = \mathbb{E}^{\mathbb{P}} \left[\nabla_\theta \log \pi_\theta(a|x) \Big|_{a=a^\theta} U(Z^\theta) \right]$$

  - Coherent risk measures [TCGM15]

$$\nabla_\theta \rho(Z^\theta) = \mathbb{E}^{\mathbb{P}^*} \left[\nabla_\theta \log \pi_\theta(a|x) \Big|_{a=a^\theta} (Z^\theta - \lambda^*) \right]$$

Proposition (Inner Gradient Formula.)

Let X_c^ϕ denote the version of X^ϕ which makes (X^θ, X_c^ϕ) comonotonic. If g is left-differentiable, then

$$\nabla_\theta L[\theta,\phi] = \mathbb{E}\Bigg[\Big(U'\left(X^\theta\right)\gamma\left(F_\theta(X^\theta)\right) \\ - p\,\Lambda\,|\Delta X_c^{\theta,\phi}|^{p-1}\,\mathrm{sgn}(\Delta X_c^{\theta,\phi})\Big)\frac{\nabla_\theta F_\theta(x)|_{x=X^\theta}}{f_\theta(X^\theta)}\Bigg] \quad (1)$$

- $\gamma\colon (0,1)\to \mathrm{R}_+$ *is given by* $\gamma(u) := \partial_- g(x)|_{x=1-u}$
- $\Lambda := (\lambda + \mu\, c[X^\theta, X^\phi]^p)\mathbb{1}_{d_p[X^\theta,X^\phi]>\varepsilon}$
- $\Delta X_c^{\theta,\phi} = X^\theta - X_c^\phi$
- *F_θ and f_θ are the cdf and pdf of X^θ*
- *G_ϕ and g_ϕ are the cdf and pdf of X^ϕ.*

Proposition (Outer Gradient Formula.)

Let X^ϕ_c denote the version of X^ϕ which makes (X^θ, X^ϕ_c) comonotonic. If g is left-differentiable, then

$$\nabla_\phi L[\theta,\phi] = \mathbb{E}\left[U'(X^\theta)\,\gamma(F_\theta(X^\theta))\,\frac{\nabla_\phi F_\theta(x)|_{x=X^\theta}}{f_\theta(X^\theta)} \right.$$
$$\left. -p\,\Lambda\,|\Delta X^{\theta,\phi}_c|^{p-1}\,\mathrm{sgn}(\Delta X^{\theta,\phi}_c)\frac{\nabla_\phi G_\phi(x)|_{x=X^\phi}}{g_\phi(X^\phi)}\right] \quad (2)$$

- $\gamma\colon (0,1) \to \mathrm{R}_+$ *is given by* $\gamma(u) := \partial_- g(x)|_{x=1-u}$
- $\Lambda := (\lambda + \mu\, c[X^\theta, X^\phi]^p)\mathbb{1}_{d_p[X^\theta, X^\phi] > \varepsilon}$
- $\Delta X^{\theta,\phi}_c = X^\theta - X^\phi_c$
- F_θ *and* f_θ *are the cdf and pdf of* X^θ
- G_ϕ *and* g_ϕ *are the cdf and pdf of* X^ϕ.

Gradients

- Gradient Formulae require estimates of f_θ, g_ϕ and $\nabla_\theta F_\theta, \nabla_\phi F_\phi$
- Use kernel density approximations (KDE) to write, e.g., from a mini-batch of data $\{(x_\theta^{(1)}, x_\phi^{(1)}), \ldots, (x_\theta^{(N)}, x_\phi^{(N)})\}$,

$$F_\theta(x) = \frac{1}{N} \sum_{i=1}^{N} \Phi_h(x - x_\theta^{(i)})$$

$$f_\theta(x) = \frac{1}{N} \sum_{i=1}^{N} \Phi'_h(x - x_\theta^{(i)})$$

$$\nabla_\theta F_\theta(x) = \frac{1}{N} \sum_{i=1}^{N} \Phi'_h(x - x_\theta^{(i)})\ \nabla_\theta x_\theta^{(i)}$$

- $\nabla_\theta x_\theta^{(i)}, \nabla_\phi x_\phi^{(i)}$ computed through back-propagation

Gradients – Randomised Policies

- One may sample actions from policy distributions instead
- In this case, $a \sim \pi(a|x)$, and transitions $x_{t+1} \sim h(x|x_t, a_t)$
- In this case, we can show that, e.g.,

$$\nabla_\phi G_\phi(x) = \mathbb{E}\left[\sum_{t=0}^{T-1} \nabla_\phi \log \pi_\phi(a|x_t)\big|_{a=a_t^\phi} \, \mathbb{1}_{X^\phi \le x}\right]$$

so that using a mini-batch of data we have

$$\nabla_\phi \hat{G}_\phi(x) = \frac{1}{N}\sum_{m=1}^{N}\sum_{t=0}^{T-1} \nabla_\phi \log \pi_\phi(a|x_t^{(m)})\big|_{a=a_t^{(m)}} \, \Phi(x_T^{(m)}) - x)$$

Algorithm

initialise networks θ, ϕ;

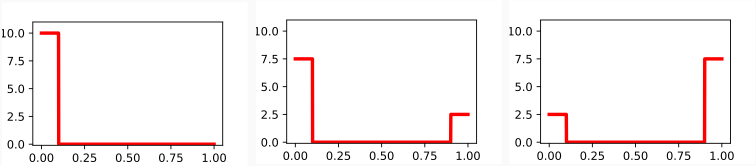
initialise Lagrangian multipliers $\lambda = 10$ and $\mu = 10$;

for $i \leftarrow 1$ **to** N **do**

- **for** $j \leftarrow 1$ **to** M_1 **do**
  - Simulate mini-batch of (X^θ, X^ϕ)
  - Estimate inner gradient $\nabla_\theta L[\theta, \phi]$ using (1);
  - Update network θ using a ADAM step;
  - Repeat until $\mathcal{R}_\gamma^U[X^\theta]$ has not improved beyond tol;
- Update multipliers: $\lambda \leftarrow \lambda + \mu\, c(\theta^*)$ and $\mu \leftarrow 2\,\mu$;
- Simulate mini-batch of (X^θ, X^ϕ);
- Estimate outer gradient $\nabla_\phi L[\theta, \phi]$ using (2);
- Update network ϕ using a ADAM step;
- Repeat until $d_p[X^\theta, X^\phi] \le \varepsilon$ and $\mathcal{R}_\gamma^U[X^\theta]$ has not improved beyond tol;

Example: $\alpha - \beta$ risk measure

- α-β risk measure $\gamma(u) = \frac{1}{\eta}\left(p\,\mathbb{1}_{\{u \leq \alpha\}} + (1-p)\,\mathbb{1}_{\{u > \beta\}}\right)$

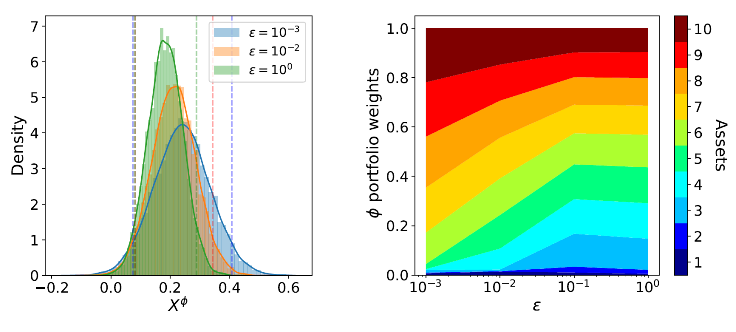


$p = 1$ $\quad$ $p > \frac{1}{2}$ $\quad$ $p < \frac{1}{2}$

- α-β risk measure is U-shaped and contains several notable special cases
  - $p = 1$ corresponds to the TVaR at level α
  - $p > \frac{1}{2}$ emphasises losses relative to gains
  - $p < \frac{1}{2}$ emphasises gains relative to losses

Example: Portfolio Optimisation

- Asset returns have idiosyncratic risk $\zeta_i \sim \mathcal{N}(i \times 3\%, i \times 2.5\%)$ and systematic risk $\psi \sim \mathcal{N}(0\%, i \times 2.5\%)$



Example: Optimising against a Benchmark

- Investor has a benchmark dynamic trading strategy ϕ
- Seek alternative strategies θ that lie within a Wasserstein ball that minimise the risk measure

Theorem ([PJ20])
The optimal quantile function is

$$g^*(u) := \left(F^{-1}_{X^\delta_T}(u) + \tfrac{1}{2\lambda_1}\left(\gamma(u) - \lambda_2\,\xi(u)\right)\right)^{\uparrow},$$

where $\lambda_1 > 0, \lambda_2 \geq 0$ chosen to satisfy constraints. Moreover, the optimal terminal wealth is

$$X^* := g^*(V).$$

Example: Optimising against a Benchmark

- Stochastic interest rate with constant elasticity of variance model

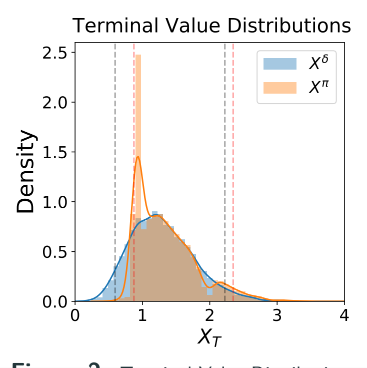


Figure 2: Terminal Value Distributions

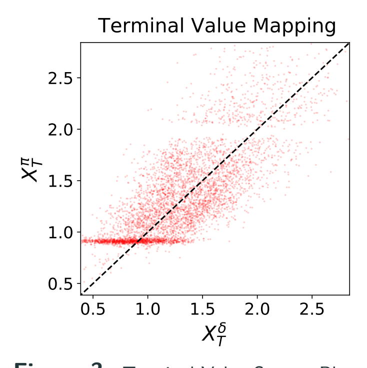
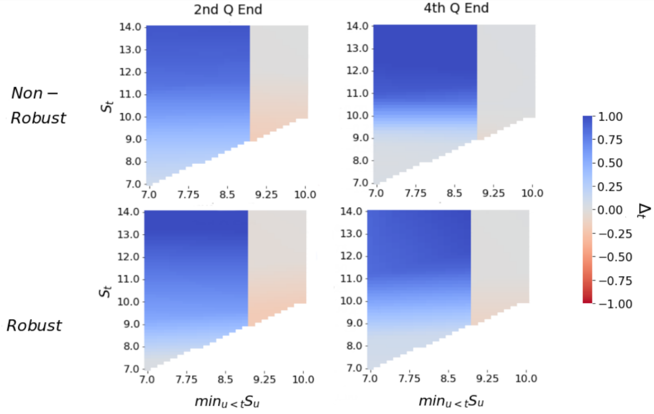


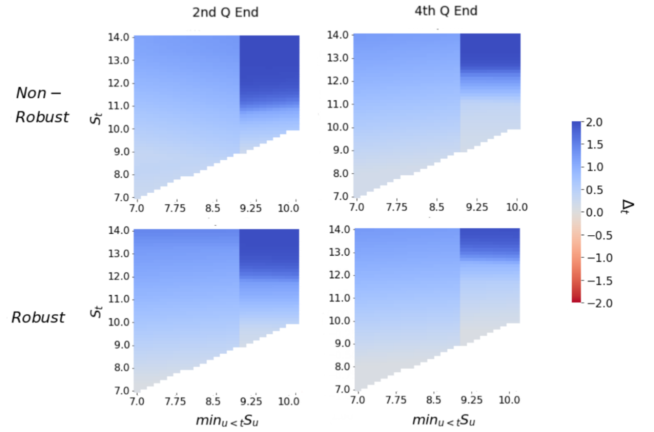
Figure 3: Terminal Value Scatter Plot

Example: Option Hedging – Down and In Call



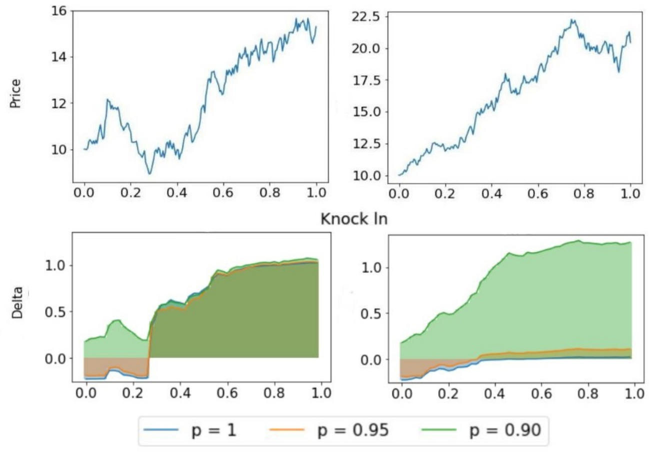
$$\mathfrak{p} = 1 - \text{CVaR}$$

Example: Option Hedging – Down and In Call



$\mathfrak{p} = 0.75$

Example: Option Hedging – Down and In Call



Various $\mathfrak{p}$ paths

Contributions & Thanks

- Developed a general formulation for Robustifying Rank Dependent Expected Utility
- Obtained explicit gradients for inner and outer problems
- Solved some interesting real-world relevant examples

code: `https://github.com/sebjai/robust-risk-aware-rl`

paper: SIAM J. Fin. Math 13(1)
`https://epubs.siam.org/doi/10.1137/21M144640X`

Thank You for Your Attention!

`http://sebastian.statistics.utoronto.ca`

References

[BM14] Ernesto G Birgin and José Mario Martínez. *Practical augmented Lagrangian methods for constrained optimization*. SIAM, 2014.

[CDJ17] Álvaro Cartea, Ryan Donnelly, and Sebastian Jaimungal. Algorithmic trading with model uncertainty. *SIAM Journal on Financial Mathematics*, 8(1):635–671, 2017.

[CO11] Alexander MG Cox and Jan Obloj. Robust hedging of double touch barrier options. *SIAM Journal on Financial Mathematics*, 2(1):141–182, 2011.

[EK18] Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1):115–166, 2018.

[LS21] Eva Lütkebohmert and Julian Sester. Robust statistical arbitrage strategies. *Quantitative Finance*, 21(3):379–402, 2021.

[PJ20] Silvana Pesenti and Sebastian Jaimungal. Portfolio optimisation within a wasserstein ball. *https://arxiv.org/abs/2012.04500*, 2020.

[PW07] Georg Pflug and David Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007.

[SMSM00] Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*, pages 1057–1063, 2000.

[TCGM15] Aviv Tamar, Yinlam Chow, Mohammad Ghavamzadeh, and Shie Mannor. Policy gradient for coherent risk measures. *Advances in Neural Information Processing Systems*, 28:1468–1476, 2015.