Risk-Averse Dynamic Programming for Markov Decision Processes

Andrzej Ruszczyński

RUTGERS

Research supported by the NSF award DMS-0603728

Outline

How to Measure Risk of Sequences?

Probability space $(\Omega, \mathcal{F}, P)$ with filtration $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T \subset \mathcal{F}$
Adapted sequence of random variables (costs) $Z_1, Z_2, \ldots, Z_T$
Spaces: $\mathcal{Z}_t = \mathcal{L}_p(\Omega, \mathcal{F}_t, P)$, $p \in [1, \infty]$, and $\mathcal{Z}_{t,T} = \mathcal{Z}_t \times \cdots \times \mathcal{Z}_T$

Conditional Risk Measure

A mapping $\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{Z}_t$ satisfying the monotonicity condition:

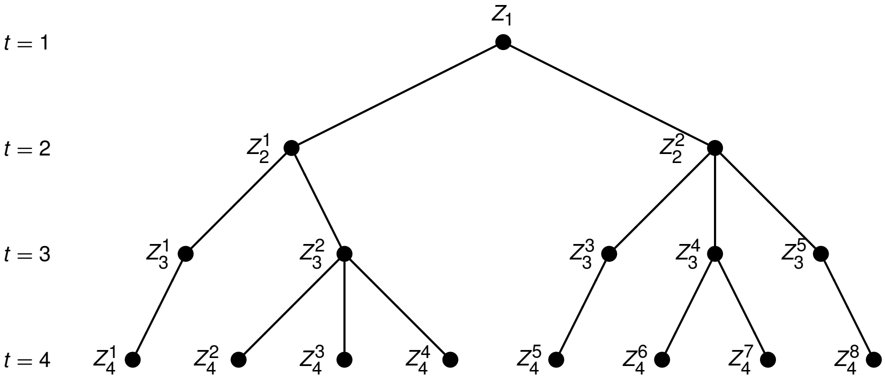
$$\rho_{t,T}(Z) \le \rho_{t,T}(W) \text{ for all } Z, W \in \mathcal{Z}_{t,T} \text{ such that } Z \le W$$

Dynamic Risk Measure

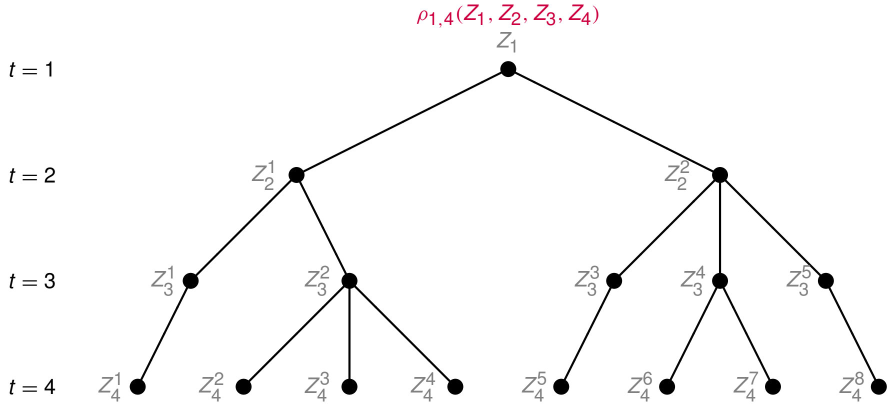
A sequence of conditional risk measures $\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{Z}_t$, $t = 1, \ldots, T$

$$\rho_{1,T}(Z_1, Z_2, Z_3, \ldots, Z_T) \in \mathcal{Z}_1 = \mathbb{R}$$
$$\rho_{2,T}(Z_2, Z_3, \ldots, Z_T) \in \mathcal{Z}_2$$
$$\rho_{3,T}(Z_3, \ldots, Z_T) \in \mathcal{Z}_3$$
$$\vdots$$

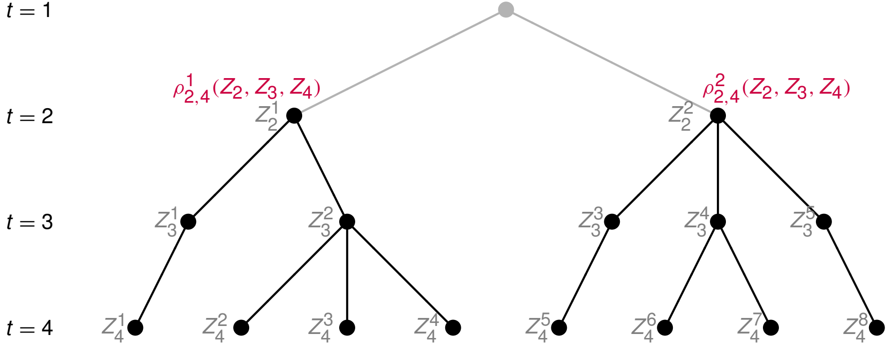
Evaluating Risk on a Scenario Tree



Evaluating Risk on a Scenario Tree



Evaluating Risk on a Scenario Tree



Time Consistency of Dynamic Risk Measures

A dynamic risk measure $\{\rho_{t,T}\}_{t=1}^{T}$ is time-consistent if for all $\tau < \theta$

$$Z_k = W_k,\ k = \tau, \dots, \theta - 1 \quad \text{and} \quad \rho_{\theta,T}(Z_\theta, \dots, Z_T) \le \rho_{\theta,T}(W_\theta, \dots, W_T)$$

imply that $\quad \rho_{\tau,T}(Z_\tau, \dots, Z_T) \le \rho_{\tau,T}(W_\tau, \dots, W_T)$

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

Risk-Averse Equivalence Theorem

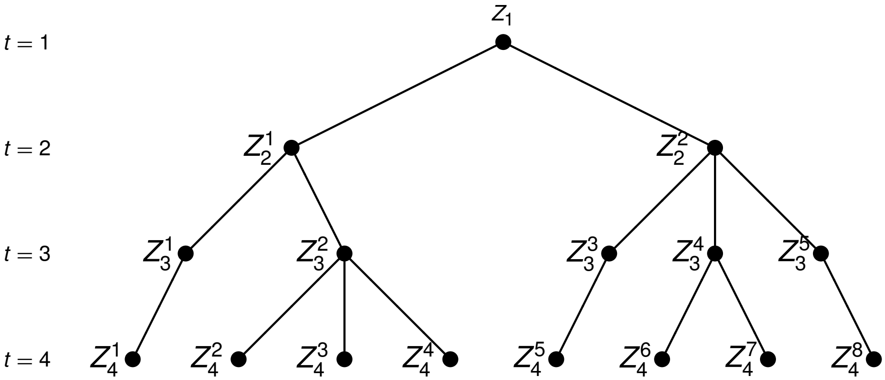
Suppose $\{\rho_{t,T}\}_{t=1}^{T}$ satisfies the conditions:

$$\rho_{t,T}(Z_t, Z_{t+1}, \dots, Z_T) = Z_t + \rho_{t,T}(0, Z_{t+1}, \dots, Z_T)$$
$$\rho_{t,T}(0, \dots, 0) = 0$$

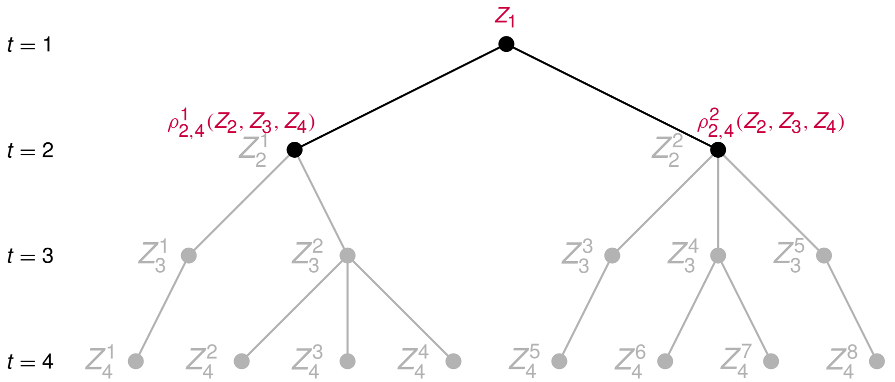
Then it is time-consistent if and only if for all $\tau \le \theta$:

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

Collapsing Subtrees by Conditional Risk Measures



Collapsing Subtrees by Conditional Risk Measures



Recursive Structure of Dynamic Risk Measures

Define one-step conditional risk measures $\rho_t : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$:

$$\rho_t(Z_{t+1}) = \rho_{t,T}(0, Z_{t+1}, 0, \dots, 0)$$

Nested Decomposition Theorem

Suppose a dynamic risk measure $\left\{\rho_{t,T}\right\}_{t=1}^{T}$ is time-consistent and

$$\rho_{t,T}(Z_t, Z_{t+1}, \dots, Z_T) = Z_t + \rho_{t,T}(0, Z_{t+1}, \dots, Z_T)$$
$$\rho_{t,T}(0, \dots, 0) = 0$$

Then for all t we have the representation

$$\rho_{t,T}(Z_t, \dots, Z_T) =$$
$$= Z_t + \rho_t\Big(Z_{t+1} + \rho_{t+1}\Big(Z_{t+2} + \dots + \rho_{T-2}\big(Z_{T-1} + \rho_{T-1}(Z_T)\big)\cdots\Big)\Big)$$

Coherent One-Step Conditional Risk Measures

Stronger assumptions about one-step measures $\rho_t : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$:

- **Convexity:** $\rho_t(\lambda Z + (1-\lambda)W) \leq \lambda \rho_t(Z) + (1-\lambda)\rho_t(W)$ $\forall\, \lambda \in (0,1),\; Z, W \in \mathcal{Z}_{t+1}$
- **Monotonicity:** If $Z \leq W$ then $\rho_t(Z) \leq \rho_t(W), \;\forall\, Z, W \in \mathcal{Z}_{t+1}$
- **Predictable Translation Equivariance:** $\rho_t(Z+W) = Z + \rho_t(W), \;\forall\, Z \in \mathcal{Z}_t, \;\; W \in \mathcal{Z}_{t+1}$
- **Positive Homogeneity:** $\rho_t(\tau Z) = \tau \rho_t(Z), \forall\, Z \in \mathcal{Z}_{t+1}, \;\; \tau \geq 0$

Scandolo ('03), Riedel ('04), R.-Shapiro ('06), Cheridito-Delbaen-Kupper ('06), Föllmer-Penner ('06), Artzner-Delbaen-Eber-Heath-Ku ('07), Pflug-Römisch ('07)

Example: Conditional Mean–Semideviation

$$\rho_t(Z_{t+1}) = \mathbb{E}[Z_{t+1}|\mathcal{F}_t] + \kappa \mathbb{E}\Big[\big(Z_{t+1} - \mathbb{E}[Z_{t+1}|\mathcal{F}_t]\big)_+^s \Big| \mathcal{F}_t\Big]^{\frac{1}{s}}$$

Here $s \in [1,p]$ and $\kappa \in [0,1]$ may be $\mathcal{F}_t$-measurable

Coherent One-Step Conditional Risk Measures

Stronger assumptions about one-step measures $\rho_t : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$:

- Convexity: $\rho_t(\lambda Z + (1-\lambda)W) \le \lambda \rho_t(Z) + (1-\lambda)\rho_t(W)$
  $\forall\, \lambda \in (0,1),\; Z, W \in \mathcal{Z}_{t+1}$
- Monotonicity: If $Z \le W$ then $\rho_t(Z) \le \rho_t(W),\; \forall\, Z, W \in \mathcal{Z}_{t+1}$
- Predictable Translation Equivariance:
  $\rho_t(Z + W) = Z + \rho_t(W),\; \forall\, Z \in \mathcal{Z}_t,\; W \in \mathcal{Z}_{t+1}$
- Positive Homogeneity: $\rho_t(\tau Z) = \tau \rho_t(Z), \forall\, Z \in \mathcal{Z}_{t+1},\; \tau \ge 0$

Scandolo ('03), Riedel ('04), R.-Shapiro ('06), Cheridito-Delbaen-Kupper ('06), Föllmer-Penner ('06), Artzner-Delbaen-Eber-Heath-Ku ('07), Pflug-Römisch ('07)

Example: Conditional Mean–Semideviation

$$\rho_t(Z_{t+1}) = \mathbb{E}[Z_{t+1}|\mathcal{F}_t] + \kappa \mathbb{E}\Big[\big(Z_{t+1} - \mathbb{E}[Z_{t+1}|\mathcal{F}_t]\big)_+^s \Big| \mathcal{F}_t\Big]^{\frac{1}{s}}$$

Here $s \in [1, p]$ and $\kappa \in [0, 1]$ may be $\mathcal{F}_t$-measurable

Coherent One-Step Conditional Risk Measures

Stronger assumptions about one-step measures $\rho_t : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$:

- Convexity: $\rho_t(\lambda Z + (1-\lambda)W) \le \lambda \rho_t(Z) + (1-\lambda)\rho_t(W)$
  $\forall\ \lambda \in (0,1),\ \ Z, W \in \mathcal{Z}_{t+1}$
- Monotonicity: If $Z \le W$ then $\rho_t(Z) \le \rho_t(W),\ \ \forall\ Z, W \in \mathcal{Z}_{t+1}$
- Predictable Translation Equivariance:
  $\rho_t(Z + W) = Z + \rho_t(W),\ \ \forall\ Z \in \mathcal{Z}_t,\ \ W \in \mathcal{Z}_{t+1}$
- Positive Homogeneity: $\rho_t(\tau Z) = \tau \rho_t(Z), \forall\ Z \in \mathcal{Z}_{t+1},\ \ \tau \ge 0$

Scandolo ('03), Riedel ('04), R.-Shapiro ('06), Cheridito-Delbaen-Kupper ('06), Föllmer-Penner ('06), Artzner-Delbaen-Eber-Heath-Ku ('07), Pflug-Römisch ('07)

Example: Conditional Mean–Semideviation

$$\rho_t(Z_{t+1}) = \mathbb{E}[Z_{t+1}|\mathcal{F}_t] + \kappa \mathbb{E}\Big[\big(Z_{t+1} - \mathbb{E}[Z_{t+1}|\mathcal{F}_t]\big)_+^s \Big| \mathcal{F}_t\Big]^{\frac{1}{s}}$$

Here $s \in [1, p]$ and $\kappa \in [0, 1]$ may be $\mathcal{F}_t$-measurable

Coherent One-Step Conditional Risk Measures

Stronger assumptions about one-step measures $\rho_t : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$:

- Convexity: $\rho_t(\lambda Z + (1-\lambda)W) \le \lambda\rho_t(Z) + (1-\lambda)\rho_t(W)$
  $\forall\, \lambda \in (0,1),\; Z, W \in \mathcal{Z}_{t+1}$
- Monotonicity: If $Z \le W$ then $\rho_t(Z) \le \rho_t(W),\; \forall\, Z, W \in \mathcal{Z}_{t+1}$
- Predictable Translation Equivariance:
  $\rho_t(Z+W) = Z + \rho_t(W),\; \forall\, Z \in \mathcal{Z}_t,\;\; W \in \mathcal{Z}_{t+1}$
- Positive Homogeneity: $\rho_t(\tau Z) = \tau\rho_t(Z), \forall\, Z \in \mathcal{Z}_{t+1},\;\; \tau \ge 0$

Scandolo ('03), Riedel ('04), R.-Shapiro ('06), Cheridito-Delbaen-Kupper ('06), Föllmer-Penner ('06), Artzner-Delbaen-Eber-Heath-Ku ('07), Pflug-Römisch ('07)

Example: Conditional Mean–Semideviation

$$\rho_t(Z_{t+1}) = \mathbb{E}[Z_{t+1}|\mathcal{F}_t] + \kappa\mathbb{E}\Big[\big(Z_{t+1} - \mathbb{E}[Z_{t+1}|\mathcal{F}_t]\big)_+^s \Big| \mathcal{F}_t\Big]^{\frac{1}{s}}$$

Here $s \in [1,p]$ and $\kappa \in [0,1]$ may be $\mathcal{F}_t$-measurable

Coherent One-Step Conditional Risk Measures

Stronger assumptions about one-step measures $\rho_t : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$:

- Convexity: $\rho_t(\lambda Z + (1-\lambda)W) \le \lambda \rho_t(Z) + (1-\lambda)\rho_t(W)$
  $\forall\, \lambda \in (0,1),\ Z, W \in \mathcal{Z}_{t+1}$
- Monotonicity: If $Z \le W$ then $\rho_t(Z) \le \rho_t(W),\ \forall\, Z, W \in \mathcal{Z}_{t+1}$
- Predictable Translation Equivariance:
  $\rho_t(Z + W) = Z + \rho_t(W),\ \forall\, Z \in \mathcal{Z}_t,\ \ W \in \mathcal{Z}_{t+1}$
- Positive Homogeneity: $\rho_t(\tau Z) = \tau \rho_t(Z), \forall\, Z \in \mathcal{Z}_{t+1},\ \ \tau \ge 0$

Scandolo ('03), Riedel ('04), R.-Shapiro ('06), Cheridito-Delbaen-Kupper ('06), Föllmer-Penner ('06), Artzner-Delbaen-Eber-Heath-Ku ('07), Pflug-Römisch ('07)

Example: Conditional Mean–Semideviation

$$\rho_t(Z_{t+1}) = \mathbb{E}[Z_{t+1}|\mathcal{F}_t] + \kappa \mathbb{E}\Big[\big(Z_{t+1} - \mathbb{E}[Z_{t+1}|\mathcal{F}_t]\big)_+^s \Big| \mathcal{F}_t\Big]^{\frac{1}{s}}$$

Here $s \in [1,p]$ and $\kappa \in [0,1]$ may be $\mathcal{F}_t$-measurable

Coherent One-Step Conditional Risk Measures

Stronger assumptions about one-step measures $\rho_t : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$:

- Convexity: $\rho_t(\lambda Z + (1-\lambda)W) \leq \lambda\rho_t(Z) + (1-\lambda)\rho_t(W)$
  $\forall\, \lambda \in (0,1),\ Z, W \in \mathcal{Z}_{t+1}$
- Monotonicity: If $Z \leq W$ then $\rho_t(Z) \leq \rho_t(W)$, $\forall\, Z, W \in \mathcal{Z}_{t+1}$
- Predictable Translation Equivariance:
  $\rho_t(Z + W) = Z + \rho_t(W),\ \forall\, Z \in \mathcal{Z}_t,\ W \in \mathcal{Z}_{t+1}$
- Positive Homogeneity: $\rho_t(\tau Z) = \tau\rho_t(Z), \forall\, Z \in \mathcal{Z}_{t+1},\ \tau \geq 0$

Scandolo ('03), Riedel ('04), R.-Shapiro ('06), Cheridito-Delbaen-Kupper ('06), Föllmer-Penner ('06), Artzner-Delbaen-Eber-Heath-Ku ('07), Pflug-Römisch ('07)

Example: Conditional Mean–Semideviation

$$\rho_t(Z_{t+1}) = \mathbb{E}[Z_{t+1}|\mathcal{F}_t] + \kappa\mathbb{E}\Big[\big(Z_{t+1} - \mathbb{E}[Z_{t+1}|\mathcal{F}_t]\big)_+^s \Big|\mathcal{F}_t\Big]^{\frac{1}{s}}$$

Here $s \in [1, p]$ and $\kappa \in [0, 1]$ may be $\mathcal{F}_t$-measurable

Multistage Risk-Averse Optimization Problems

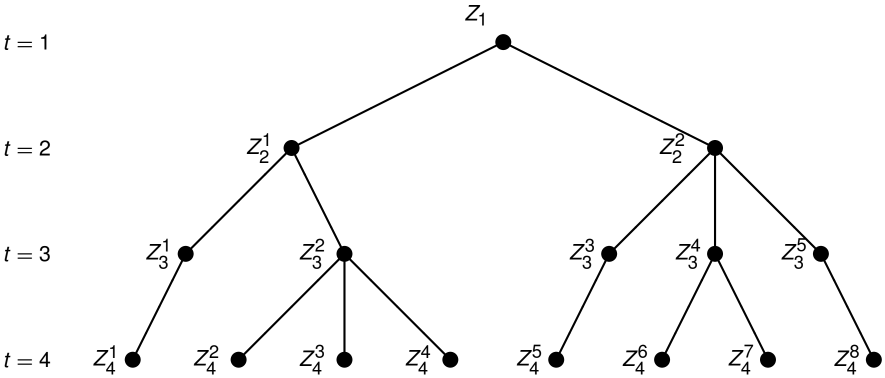
Probability Space: $(\Omega, \mathcal{F}, P)$ with filtration $\mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T \subset \mathcal{F}$
Decision Variables: $x_t(\omega)$, $\omega \in \Omega$, $t = 1, \dots, T$
Nonanticipativity: Each x_t is $\mathcal{F}_t$-measurable
Cost per Stage: $Z_t(x_t)$ with realizations $Z_t(x_t(\omega), \omega)$, $\omega \in \Omega$
Objective Function: Time-consistent dynamic measure of risk

Interchangeability Principle

$$\min_{x_1, x_2(\cdot), \dots, x_T(\cdot)} \Big\{ Z_1(x_1) + \rho_1\Big(Z_2(x_2) + \rho_2\Big(Z_3(x_3) + \dots \\ \cdots + \rho_{T-2}\big(Z_{T-1}(x_{T_1} + \rho_{T-1}(Z_T(x_T))\big) \cdots \Big)\Big)\Big\}$$

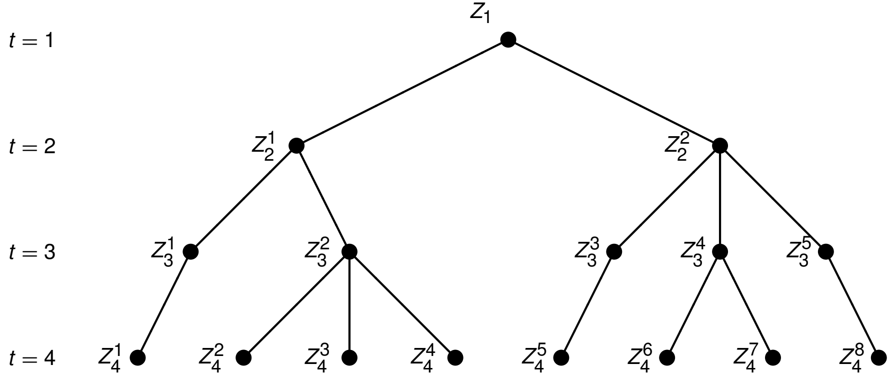
$$= \min_{x_1} \Big\{ Z_1(x_1) + \rho_1\Big[\min_{x_2}\Big(Z_2(x_2) + \rho_2\Big[\min_{x_3}\Big(Z_3(x_3) + \dots \\ \cdots + \rho_{T-2}\big[\min_{x_{T-1}}\big(Z_{T-1}(x_{T_1}) + \rho_{T-1}(\min_{x_T} Z_T(x_T))\big)\big] \cdots \Big)\Big]\Big)\Big]\Big\}$$

Interchangeability on a Scenario Tree

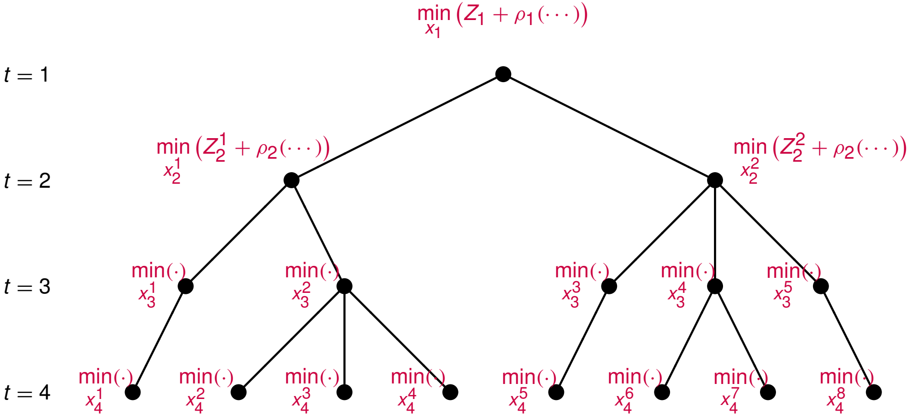


Interchangeability on a Scenario Tree

$$\min_{x_1, x_2(\cdot), x_3(\cdot), x_4(\cdot)} \rho_{1,4}(Z_1, Z_2, Z_3, Z_4)$$



Interchangeability on a Scenario Tree



Controlled Markov Models

- State space $\mathcal{X}$ (Polish with Borel σ-algebra)
- Control space $\mathcal{U}$ (Polish with Borel σ-algebra)
- Feasible control sets $U_t : \mathcal{X} \rightrightarrows \mathcal{U}$, $t = 1, 2, \dots$
- Controlled transition kernels $Q_t : \text{graph}(U_t) \to \mathcal{P}$, $t = 1, 2, \dots$
  $\mathcal{P}$ - set of probability measures on $\mathcal{X}$
- Cost functions $c_t : \text{graph}(U_t) \to \mathbb{R}$, $t = 1, 2, \dots$
- State history $\mathcal{X}^t$ (up to time $t = 1, 2, \dots$)
- Policy $\pi_t : \mathcal{X}^t \to \mathcal{U}$, $t = 1, 2, \dots$ (always with values in $U_t(x_t)$)
- Markov policy $\pi_t : \mathcal{X} \to \mathcal{U}$, $t = 1, 2, \dots$
  (stationary if $\pi_t = \pi_1$ for all t)

$$x_t \longrightarrow u_t = \pi_t(x_t)$$

$$(x_t, u_t) \longrightarrow x_{t+1} \sim Q_t(x_t, u_t)$$

Two Basic Risk-Neutral Control Problems

Finite horizon expected cost problem:

$$\min_{\pi_1,\dots,\pi_T} \mathbb{E}\left[\sum_{t=1}^{T} c_t(x_t, u_t) + c_{T+1}(x_{T+1})\right]$$

with controls $u_t = \pi_t(x_1, \dots, x_t)$

Infinite horizon discounted expected cost problem:

$$\min_{\pi_1,\pi_2,\dots} \mathbb{E}\left[\sum_{t=1}^{\infty} \alpha^{t-1} c_t(x_t, u_t)\right]$$

- Both problems have optimal solutions in form of Markov policies
- Optimal policies can be found by dynamic programming equations

Our Intention

Introduce risk aversion to both problems by replacing the expected value by dynamic risk measures

Using Dynamic Risk Measures for Markov Decision Processes

- Controlled Markov process x_t, $t = 1, \ldots, T, T+1$
- Policy $\Pi = \{\pi_1, \pi_2, \ldots, \pi_T\}$ defines $u_t = \pi_t(x_t)$
- Cost sequence $c_t(x_t, u_t)$, $t = 1, \ldots, T$, and $c_{T+1}(x_{T+1})$
- Dynamic time-consistent risk measure

$$J(\Pi) = c_1(x_1, u_1) + \rho_1\Big(c_2(x_2, u_2) + \rho_2\Big(c_3(x_3, u_3) + \cdots + \rho_{T-1}\big(c_T(x_T, u_T) + \rho_T(c_{T+1}(x_{T+1}))\big)\cdots\Big)\Big)$$

- Risk-averse optimal control problem

$$\min_{\Pi} J(\Pi)$$

Difficulty

The value of $\rho_t(\cdot)$ is $\mathcal{F}_t$-measurable and is allowed to depend on the entire history of the process. We cannot expect a Markov optimal policy if our attitude to risk depends on the whole past

New Construction of a Conditional Risk Measure

- $\mathcal{B}$ - Borel σ-field on $\mathcal{X}$, P_0 - probability measure on $(\mathcal{X}, \mathcal{B})$
- Spaces: $\mathcal{V} = \mathcal{L}_p(\mathcal{X}, \mathcal{B}, P_0)$, $\mathcal{Y} = \mathcal{L}_q(\mathcal{X}, \mathcal{B}, P_0)$ $(\frac{1}{p} + \frac{1}{q} = 1)$
- Densities on $(\mathcal{X}, \mathcal{B})$

$$\mathcal{M} = \left\{ m \in \mathcal{Y} : \int_{\mathcal{X}} m(x)\, P_0(dx) = 1,\ m \geq 0 \right\}$$

- Pairing of the spaces $\mathcal{V}$ and $\mathcal{Y}$ with the bilinear form

$$\langle v, m \rangle = \int_{\mathcal{X}} v(x) m(x)\ P_0(dx)$$

Risk Transition Mapping Associated with a Kernel $Q : \text{graph}(U) \to \mathcal{M}$

A measurable functional $\sigma : \mathcal{V} \times \mathcal{X} \times \mathcal{M} \to \mathbb{R}$ satisfying for every measurable selection $u(\cdot)$ of $U(\cdot)$ the conditions

(i) For every $x \in \mathcal{X}$ the functional $v \mapsto \sigma\big(v, x, Q(x, u(x))\big)$ is a coherent measure of risk on $\mathcal{V}$

(ii) For every $v \in \mathcal{V}$ the function $x \mapsto \sigma\big(v, x, Q(x, u(x))\big)$ is in $\mathcal{V}$

Dual Representation of Risk Transition Mappings

If the mapping $\sigma(v, x, m)$ is lower semicontinuous with respect to v, then there exist convex sets $\mathcal{A}(x, m)$ such that

$$\sigma(v, x, m) = \sup_{\mu \in \mathcal{A}(x,m)} \langle v, \mu \rangle$$

Example: Mean–Semideviation Mapping

$$\sigma(v, x, m) = \langle v, m \rangle + \kappa(x) \Big(\big\langle (v - \langle v, m \rangle)_+^s, m \big\rangle \Big)^{\frac{1}{s}}$$

For $s > 1$ we obtain

$$\mathcal{A}(x, m) = \left\{ g = m\big(1 + h - \langle h, m \rangle\big) : \Big(\big\langle |h|^{\frac{s}{s-1}}, m \big\rangle \Big)^{\frac{s-1}{s}} \le \kappa(x), \;\; h \ge 0 \right\}$$

and for $s = 1$ we have

$$\mathcal{A}(x, m) = \left\{ g = m\big(1 + h - \langle h, m \rangle\big) : \sup_{y \in \mathcal{X}} |h(y)| \le \kappa(x), \;\; h \ge 0 \right\}$$

Markov Risk Measures

Assumption: The controlled kernels Q_t have values in the set $\mathcal{M}$ (with densities with respect to P_0)

A one-step conditional risk measure $\rho_t : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$ is a Markov risk measure with respect to the controlled Markov process $\{x_t\}$, if there exists a risk transition mapping $\sigma_t : \mathcal{V} \times \mathcal{X} \times \mathcal{M} \to \mathbb{R}$ such that for all $v \in \mathcal{V}$ and for all measurable $u_t \in U_t(x_t)$ we have

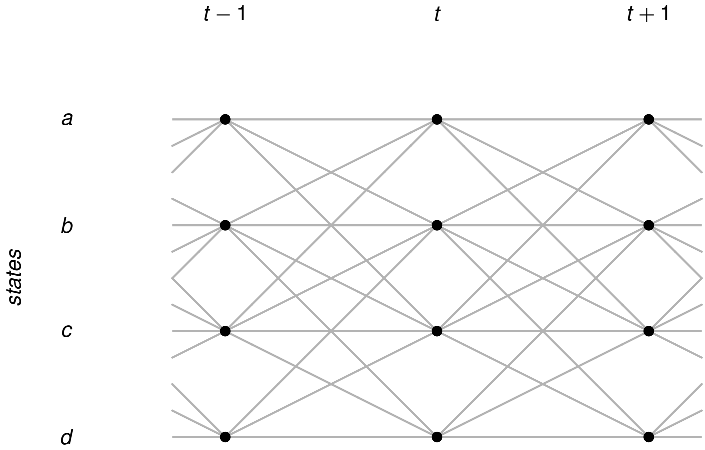
$$\rho_t(v(x_{t+1})) = \sigma_t\big(v, x_t, Q_t(x_t, u_t)\big)$$

Duality: $\rho_t(v(x_{t+1})) = \sup\limits_{\mu \in \mathcal{A}_t(x_t, Q_t(x_t, u_t))} \langle v, \mu \rangle$

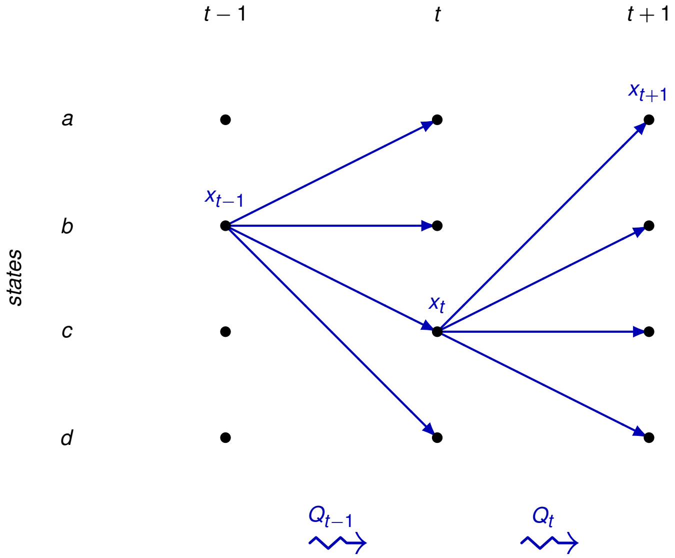
$\mathcal{A}_t(x_t, Q_t(x_t, u_t))$ – controlled multikernel

In the risk neutral setting, when $\rho_t(v(x_{t+1})) = \mathbb{E}[v(x_{t+1}) | \mathcal{F}_t]$ we have a single-valued controlled kernel $\mathcal{A}_t(x_t, Q_t(x_t, u_t)) = \{Q_t(x_t, u_t)\}$.
Risk-averse preferences $\leftrightarrows$ Ambiguity in the transition kernel

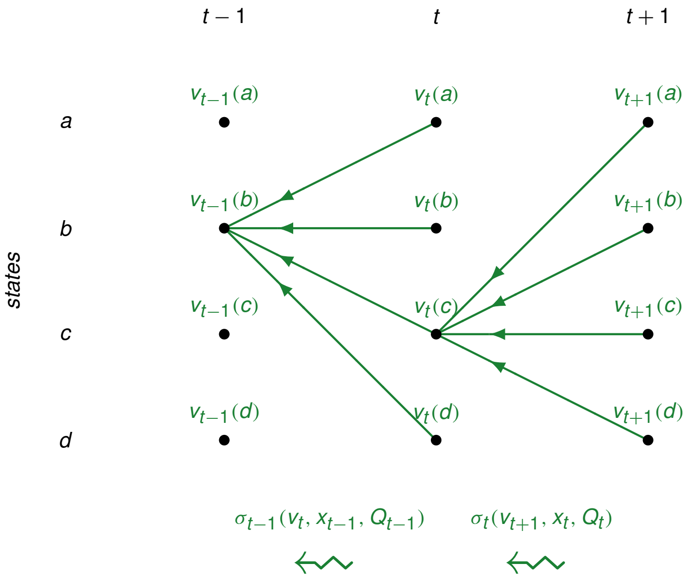
Markov Risk Evaluation



Markov Risk Evaluation



Markov Risk Evaluation



Finite Horizon Risk-Averse Control Problem

Consider a controlled Markov process $\{x_t\}$ with $u_t = \pi_t(x_1, \dots, x_t)$.

Risk-averse optimal control problem:

$$\min_{\Pi} \; c_1(x_1, u_1) + \rho_1\Big(c_2(x_2, u_2) + \rho_2\Big(c_3(x_3, u_3) + \cdots$$
$$\cdots + \rho_{T-1}\big(c_T(x_T, u_T) + \rho_T(c_{T+1}(x_{T+1}))\big) \cdots \Big)\Big)$$

Theorem

If the conditional measures ρ_t are Markov (+ technical conditions), then the optimal solution is given by the dynamic programming equations:

$$v_{T+1}(x) = c_{T+1}(x), \quad x \in \mathcal{X}$$
$$v_t(x) = \min_{u \in U_t(x)} \Big\{ c_t(x, u) + \sigma_t\big(v_{t+1}, x, Q_t(x, u)\big) \Big\}, \quad t = T, \dots, 1$$

Optimal Markov policy $\hat{\Pi} = \{\hat{\pi}_1, \dots, \hat{\pi}_T\}$ - the minimizers above

Finite Horizon Risk-Averse Control Problem

Consider a controlled Markov process $\{x_t\}$ with $u_t = \pi_t(x_1, \dots, x_t)$.

Risk-averse optimal control problem:

$$\min_{\Pi} \; c_1(x_1, u_1) + \rho_1\Bigg(c_2(x_2, u_2) + \rho_2\bigg(c_3(x_3, u_3) + \cdots \\ \cdots + \rho_{T-1}\big(c_T(x_T, u_T) + \rho_T(c_{T+1}(x_{T+1}))\big) \cdots \bigg)\Bigg)$$

Theorem

If the conditional measures ρ_t are Markov (+ technical conditions), then the optimal solution is given by the dynamic programming equations:

$$v_{T+1}(x) = c_{T+1}(x), \quad x \in \mathcal{X}$$

$$v_t(x) = \min_{u \in U_t(x)} \left\{ c_t(x,u) + \sup_{\mu \in \mathcal{A}_t(x, Q_t(x,u))} \langle v_{t+1}, \mu \rangle \right\}, \quad t = T, \dots, 1$$

Optimal Markov policy $\hat{\Pi} = \{\hat{\pi}_1, \dots, \hat{\pi}_T\}$ - the minimizers above

Discounted Risk Measures for Infinite Sequences

- $\{\mathcal{F}_t\}$ - filtration on $(\Omega, \mathcal{F})$
- Z_t, $t = 1, 2 \dots$ - adapted sequence of random variables
- $\mathcal{Z}_t = \mathcal{L}_p(\Omega, \mathcal{F}_t, P)$, $\mathcal{Z} = \mathcal{Z}_1 \times \mathcal{Z}_2 \times \cdots$
- $\rho_t : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$ - conditional risk mappings

Fix the discount factor $\alpha \in (0, 1)$. For $T = 1, 2, \dots$ define

$$\begin{aligned}\rho^{\alpha}_{1,T}(Z_1, Z_2, \dots, Z_T) &= \rho_{1,T}(Z_1, \alpha Z_2, \dots, \alpha^{T-1} Z_T) \\ &= Z_1 + \rho_1\Big(\alpha Z_2 + \rho_2\big(\alpha^2 Z_3 + \cdots + \rho_{T-1}(\alpha^{T-1} Z_T) \cdots\big)\Big)\end{aligned}$$

Discounted Risk Measure

$$\varrho^{\alpha}(Z) = \lim_{T \to \infty} \rho^{\alpha}_{1,T}(Z_1, Z_2, \dots, Z_T)$$

It is well defined, convex, monotone, and positively homogeneous, whenever $\max_t \operatorname{ess\,sup} |Z_t(\omega)| < \infty$

Discounted Infinite Horizon Problem

We consider a controlled stationary Markov process $\{x_t\}$, $t = 1, 2, \dots$ with a discounted measure of risk ($0 < \alpha < 1$):

$$\min_{\Pi} J(\Pi, x_1) = \varrho^{\alpha}\big(c(x_1, u_1), c(x_2, u_2), \cdots\big)$$
$$= c(x_1, u_1) + \rho_1\Big(\alpha c(x_2, u_2) + \rho_2\big(\alpha^2 c(x_3, u_3) + \cdots\big)\Big)$$

Conditional Markov risk measures ρ_t are stationary, if they share the same risk transition mapping $\sigma : \mathcal{X} \times \mathcal{V} \times \mathcal{M} \to \mathbb{R}$

Theorem

If the conditional measures ρ_t are Markov and stationary, then the optimal value function $\hat{v}(x)$ satisfies the dynamic programming equation:

$$v(x) = \min_{u \in U(x)} \Big\{c(x, u) + \alpha\, \sigma\big(v, x, Q(x, u)\big)\Big\}, \quad x \in \mathcal{X}$$

Optimal stationary Markov policy $\hat{\Pi} = \{\hat{\pi}, \hat{\pi}, \dots\}$ - the minimizer above

Discounted Infinite Horizon Problem

We consider a controlled stationary Markov process $\{x_t\}$, $t = 1, 2, \ldots$ with a discounted measure of risk ($0 < \alpha < 1$):

$$\min_{\Pi} J(\Pi, x_1) = \varrho^{\alpha}\big(c(x_1, u_1), c(x_2, u_2), \cdots\big)$$
$$= c(x_1, u_1) + \rho_1\Big(\alpha c(x_2, u_2) + \rho_2\big(\alpha^2 c(x_3, u_3) + \cdots\big)\Big)$$

Conditional Markov risk measures ρ_t are stationary, if they share the same risk transition mapping $\sigma : \mathcal{X} \times \mathcal{V} \times \mathcal{M} \to \mathbb{R}$

Theorem

If the conditional measures ρ_t are Markov and stationary, then the optimal value function $\hat{v}(x)$ satisfies the dynamic programming equation:

$$v(x) = \min_{u \in U(x)} \Big\{ c(x, u) + \alpha \sup_{\mu \in \mathcal{A}(x, Q(x,u))} \langle v, \mu \rangle \Big\}, \quad x \in \mathcal{X}$$

Optimal stationary Markov policy $\hat{\Pi} = \{\hat{\pi}, \hat{\pi}, \ldots\}$ - the minimizer above

Value Iteration

Dynamic programming equation:

$$v(x) = \min_{u \in U(x)} \left\{ c(x,u) + \alpha\, \sigma\big(v, x, Q(x,u)\big) \right\}, \quad x \in \mathcal{X}$$

Observation: The operator on the right hand side is monotone and is a contraction in $\mathcal{L}_\infty(\mathcal{X}, \mathcal{B}, P_0)$ for $\alpha \in (0,1)$

Theorem

The sequence $\{v^k\}$ generated by the value iteration method

$$v^{k+1}(x) = \min_{u \in U(x)} \left\{ c(x,u) + \alpha\, \sigma\big(v^k, x, Q(x,u)\big) \right\}, \quad x \in \mathcal{X}, \quad k = 1, 2, \ldots$$

is convergent linearly in $\mathcal{L}_\infty(\mathcal{X}, \mathcal{B}, P_0)$ to the optimal value function $\hat{v}$, with quotient α. If $v^1 = 0$, then the sequence $\{v^k\}$ is nondecreasing

Policy Iteration

- For $k = 0, 1, 2, \ldots$, given a stationary Markov policy $\{\pi^k, \pi^k, \ldots\}$, find the value function v^k by solving the nonsmooth equation

$$v(x) = c(x, \pi^k(x)) + \alpha\sigma\big(v, x, Q(x, \pi^k(x))\big), \quad x \in \mathcal{X}$$

- Find the next policy $\pi^{k+1}(\cdot)$ by one-step optimization

$$\pi^{k+1}(x) = \operatorname*{argmin}_{u \in U(x)} \Big\{ c(x, u) + \alpha\sigma\big(v^k, x, Q(x, u)\big) \Big\}, \quad x \in \mathcal{X}$$

- Increase k by 1, and continue.

Theorem

The sequence of functions v^k, $k = 1, 2, \ldots$, is nonincreasing and convergent to the unique bounded solution $\hat{v}(\cdot)$ of the dynamic programming equation

Specialized Nonsmooth Newton Method

The nonsmooth equation at each step of policy iteration

$$v(x) = \bar{c}(x) + \alpha \sup_{\mu \in \bar{\mathcal{A}}(x)} \langle v, \mu \rangle, \quad x \in \mathcal{X}$$

with $\bar{c}(x) = c(x, \pi^k(x))$ and $\bar{\mathcal{A}}(x) = \mathcal{A}(x, Q(x, \pi^k(x)))$

- For $\ell = 1, 2, \ldots,$ having an approximate value function v_ℓ calculate

  the kernel $\quad \mu_\ell(x) = \operatorname*{argmax}_{\mu \in \bar{\mathcal{A}}(x)} \langle v_\ell, \mu \rangle, \quad x \in \mathcal{X}$

- Find $v_{\ell+1}$ by solving the linear equation

$$v(x) = \bar{c}(x) + \alpha \langle v, \mu_\ell(x) \rangle, \quad x \in \mathcal{X}$$

- Increase ℓ by one, and continue.

Theorem

For every initial function v_1 the sequence $\{v_\ell\}$ generated by the Newton method is convergent to the unique solution v^* of the policy equation. Moreover, the sequence is monotone.