Risk-Averse Control of Discrete-Time Partially Observable Systems

Andrzej Ruszczyński

RUTGERS

SESO Paris

June, 2016

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: $\mathcal{Z}_t = \mathcal{L}_{\bar{s}}(\Omega, \mathcal{F}_t, P)$, $\bar{s} \in [1, \infty)$, 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 $\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{Z}_t$, $t = 1, \ldots, T$

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

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

Nested Decomposition Theorem

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

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

Then for all t we have the representation

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

History-Dependent Controlled Discrete-Time Models

- State space $\mathcal{X}$ (Borel)
- Control space $\mathcal{U}$ (Borel)
- State-control histories

  $$g_t = (x_1, u_1, \dots, x_{t-1}, u_{t-1}, x_t) \in \mathcal{G}_t = (\mathcal{X} \times \mathcal{U})^{t-1} \times \mathcal{X}$$

- Feasible control sets $U_t : \mathcal{G}_t \rightrightarrows \mathcal{U}$, $t = 1, 2, \dots$
- Controlled transition kernels $Q_t : \text{graph}(U_t) \to \mathcal{P}(\mathcal{X})$, $t = 1, 2, \dots$

  $\mathcal{P}(\mathcal{X})$ - set of probability measures on $\mathcal{X}$
- Cost functions $c_t : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$, $t = 1, 2, \dots$
- State history $h_t = (x_1, \dots, x_t) \in \mathcal{X}^t$ (up to time $t = 1, 2, \dots$)
- Policy $\pi_t : \mathcal{X}^t \to \mathcal{U}$, $t = 1, 2, \dots$

For any policy π, the transition kernels can be rewritten as measurable functions from $\mathcal{X}^t$ to $\mathcal{P}(\mathcal{X})$:

$$Q_t^{\pi} : (x_1, \dots, x_t) \mapsto Q_t\big(x_1, \pi_1(x_1), \dots, x_t, \pi_t(x_1, \dots, x_t)\big)$$

Risk-Neutral Total Cost Problem

Finite horizon expected cost problem:

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

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

Standard Results for Markov systems:

- A deterministic Markov policy is optimal
- Optimal policy can be found by dynamic programming equations
- In a partially observable case, history dependence can be handled via belief state and Bayes operator to reduce to Markov model

Our Intention

Introduce risk aversion to the problem by replacing the expected value by dynamic risk measures

Using Dynamic Risk Measures for Controlled Processes

- Controlled process x_t^π, $t = 1, \ldots, T$
- Policy $\pi = \{\pi_1, \ldots, \pi_T\}$ with $u_t = \pi_t(x_1, \ldots, x_t)$ implies measure P^π
- Cost sequence $Z_t^\pi = c(x_t, u_t)$ (bounded), $t = 1, \ldots, T$,
- Dynamic time-consistent risk measure

$$J_T(\pi) = Z_1^\pi + \rho_1^\pi\Big(Z_2^\pi + \cdots + \rho_{T-1}^\pi(Z_T^\pi)\cdots\Big)$$

- Risk-averse optimal control problem: $\min_\pi J_T(\pi)$

Difficulties

- Probability measure P^π, processes x_t^π and Z_t^π depend on policy π
- Risk measures $\rho_t^\pi(\cdot)$ depend on π and history;
  no Markov policies for Markov models

Idea

We only need to measure risk of random sequences that may occur

History $h_t = (x_1, \ldots, x_t)$. Process $Z_t^\pi(h_t) = c(x_t, \pi_t(h_t))$, $t = 1, \ldots, T$

A family of conditional risk measures $\{\rho_{t,T}^\pi\}_{t=1,\ldots,T}^{\pi\in\Pi}$ is stochastically conditionally time-consistent if for all feasible policies π, π', all $1 \le t \le T-1$, and for all histories $h_t \in \mathcal{X}^t$, the relations

$$Z_t^\pi(h_t) = Z_t^{\pi'}(h_t)$$

$$\big(\rho_{t+1,T}^\pi(Z_{t+1}^\pi, \ldots, Z_T^\pi)\big|H_t^\pi = h_t\big) \preceq_{\text{st}} \big(\rho_{t+1,T}^{\pi'}(Z_{t+1}^{\pi'}, \ldots, Z_T^{\pi'})\big|H_t^{\pi'} = h_t\big)$$

imply

$$\rho_{t,T}^\pi(Z_t^\pi, \ldots, Z_T^\pi)(h_t) \le \rho_{t,T}^{\pi'}(Z_t^{\pi'}, \ldots, Z_T^{\pi'})(h_t)$$

The conditional stochastic order $\preceq_{\text{st}}$:

$$Q_t^\pi(h_t)\big(\{y : Z_t^\pi(h_t) + \rho_{t+1,T}^\pi(Z_{t+1}^\pi, \ldots, Z_T^\pi)(h_t, y) > \eta\}\big)$$
$$\le Q_t^{\pi'}(h_t)\big(\{y : Z_t^{\pi'}(h_t) + \rho_{t+1,T}^{\pi'}(Z_{t+1}^{\pi'}, \ldots, Z_T^{\pi'})(h_t, y) > \eta\}\big)$$

A family of process-based dynamic risk measures $\{\rho^{\pi}_{t,T}\}^{\pi\in\Pi}_{t=1,\dots,T}$ for a Markov decision problem is Markovian if for all Markov policies $\pi \in \Pi$, for any measurable and bounded $c_1,\dots,c_T : \mathcal{X}\times\mathcal{U}\to\mathbb{R}$, and for all $h_t = (x_1,\dots,x_t)$ and $h'_t = (x'_1,\dots,x'_t)$ such that $x_t = x'_t$, we have

$$\rho^{\pi}_{t,T}\big(c_t(X_t,\pi_t(X_t)),\dots,c_T(X_T,\pi_T(X_T))\big)(h_t) = \rho^{\pi}_{t,T}\big(c_t(X_t,\pi_t(X_t)),\dots,c_T(X_T,\pi_T(X_T))\big)(h'_t).$$

If the current state x_t is the same, and the same Markov policy π is used, then the risk is the same. The risk measure can be written as a function of the state:

$$\rho^{\pi}_{t,T}\big(c_t(X_t,\pi_t(X_t)),\dots,c_T(X_T,\pi_T(X_T))\big)(x_t)$$

Structure of Process-Based Risk Measures (with Jingnan Fan)

For a fixed history-dependent policy Π and every $h_t \in \mathcal{X}^t$, we write

$$v_t^\pi(h_t) = \rho_{t,T}^\pi\big(c_t(X_t, \pi_t(H_t)), \dots, c_T(X_T, \pi_T(H_T))\big)(h_t)$$

If a family of process-based dynamic risk measures $\{\rho_{t,T}^\pi\}_{t=1,\dots,T}^{\Pi\in\Pi}$ is translation-invariant and stochastically conditionally time-consistent, then there exist transition risk mappings

$$\sigma_t : \Big\{\bigcup_{\pi\in\Pi} \text{graph}(Q_t^\pi)\Big\} \times \mathcal{V} \to \mathbb{R},$$

($\mathcal{V}$ - space of measurable bounded functions on $\mathcal{X}$)
such that for all π, for all t, and all h_t, the functional $\sigma_t\big(h_t, Q_t^\pi(h_t), \cdot\big)$ is a law-invariant risk measure on $\big(\mathcal{X}, \mathcal{B}(\mathcal{X}), Q_t^\pi\big)$ and we have

$$v_t^\pi(h_t) = c_t(x_t, \pi_t(h_t)) + \sigma_t\big(h_t, Q_t^\pi(h_t), v_{t+1}^\pi(h_t, \cdot)\big), \;\; t = 1, \dots, T-1$$

Structure of Markovian Risk Measures (with Jingnan Fan)

For a fixed history-dependent policy Π and every $h_t \in \mathcal{X}^t$, we write

$$v_t^\pi(h_t) = \rho_{t,T}^\pi\Big(c_t(X_t, \pi_t(H_t)), \dots, c_T(X_T, \pi_T(H_T))\Big)(h_t)$$

If a family of process-based dynamic risk measures $\left\{\rho_{t,T}^\pi\right\}_{t=1,\dots,T}^{\pi\in\Pi}$ is Markovian, translation-invariant, and stochastically conditionally time-consistent, then there exist transition risk mappings

$$\sigma_t : \Big\{\big(x, Q_t(x,u)\big) \,:\, u \in U(x),\ x \in \mathcal{X}\Big\} \times \mathcal{V} \to \mathbb{R}, \quad t = 1, \dots, T-1$$

($\mathcal{V}$ - space of measurable bounded functions on $\mathcal{X}$)
such that for all $\pi \in \Pi$, for all $t = 1, \dots, T-1$, and all $h_t \in \mathcal{X}^t$, the functional $\sigma_t\big(x_t, Q_t(x_t, \pi_t(h_t), \cdot)\big)$ is a law-invariant risk measure on $\big(\mathcal{X}, \mathcal{B}(\mathcal{X}), Q_t\big)$ and

$$v_t^\pi(h_t) = c_t(x_t, \pi_t(h_t)) + \sigma_t\Big(x_t, Q_t(x_t, \pi_t(h_t)), v_{t+1}^\pi(h_t, \cdot)\Big), \ t = 1, \dots, T-1$$

Structure of Markovian Risk Measures (with Jingnan Fan)

For a fixed history-dependent policy Π and every $h_t \in \mathcal{X}^t$, we write

$$v_t^\pi(h_t) = \rho_{t,T}^\pi\big(c_t(X_t, \pi_t(H_t)), \ldots, c_T(X_T, \pi_T(H_T))\big)(h_t)$$

If a family of process-based dynamic risk measures $\{\rho_{t,T}^\pi\}_{t=1,\ldots,T}^{\pi\in\Pi}$ is Markovian, translation-invariant, and stochastically conditionally time-consistent, then there exist transition risk mappings

$$\sigma_t : \big\{\big(x, Q_t(x,u)\big) : u \in U(x),\ x \in \mathcal{X}\big\} \times \mathcal{V} \to \mathbb{R}, \quad t = 1, \ldots, T-1$$

($\mathcal{V}$ - space of measurable bounded functions on $\mathcal{X}$)
such that for all $\pi \in \Pi$, for all $t = 1, \ldots, T-1$, and all $h_t \in \mathcal{X}^t$, the functional $\sigma_t\big(x_t, Q_t(x_t, \pi_t(h_t), \cdot)\big)$ is a law-invariant risk measure on $\big(\mathcal{X}, \mathcal{B}(\mathcal{X}), Q_t\big)$ and (for Markovian π)

$$v_t^\pi(x_t) = c_t(x_t, \pi_t(x_t)) + \sigma_t\big(x_t, Q_t(x_t, \pi_t(x_t)), v_{t+1}^\pi(\cdot)\big),\ t = 1, \ldots, T-1$$

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:

$$\min_{\pi} \; J_T(\pi) = c_1(x_1, u_1) + \rho_1^{\pi}\Big(c_2(X_2, u_2) + \cdots + \rho_{T-1}^{\pi}\big(c_T(X_T, u_T)\big) \cdots \Big)$$

Theorem

If the conditional measures ρ_t^{π} are Markovian (+ general conditions), then the optimal solution is given by the dynamic programming equations:

$$v_T(x) = \min_{u \in U_T(x)} \; c_T(x, u), \quad x \in \mathcal{X}$$

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

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

Partially Observable Models

- Markov Process: $\{X_t, Y_t\}_{t=1,\dots,T}$ on the state space $\mathcal{X} \times \mathcal{Y}$
- The process $\{Y_t\}$ is observable, while $\{X_t\}$ is not observable
- Control sets: $U_t : \mathcal{Y} \rightrightarrows \mathcal{U}$, $t = 1, \dots, T$.
- Transition kernel:

$$\mathbb{P}[(X_{t+1}, Y_{t+1}) \in C \mid x_1, y_1, u_1, \dots, x_t, y_t, u_t] = Q_t(x_t, y_t, u_t)(C),$$

- Belief State: Conditional distribution of X_t given initial distribution ξ_1 and history $g_t = (\xi_1, y_1, u_1, y_2, \dots, u_{t-1}, y_t)$

$$\big[\Xi_t(g_t)\big](A) = \mathbb{P}[X_t \in A \mid g_t], \quad \forall\, A \in \mathcal{B}(\mathcal{X}), \quad t = 1, \dots, T$$

Conditional distribution of the observable part:

$$\mathbb{P}\left[Y_{t+1} \in B \mid g_t, u_t\right] = \int_{\mathcal{X}} \big[Q_t^Y(\cdot, y_t, u_t)\big](B)\; d\Xi_t(g_t),$$

where $Q_t^Y(x_t, y_t, u_t)$ is the marginal of $Q_t(x_t, y_t, u_t)$ on the space $\mathcal{Y}$

Transition of the belief state:

$$\Xi_{t+1}(g_{t+1}) = \Phi_t(y_t, \Xi_t(g_t), u_t, y_{t+1})$$

Bayes operator with kernels given by density functions

If $\mathcal{X} = \{x^1, \dots, x^n\}$ is finite and kernels $Q_t(x, y, u)$ have density $q_t(x', \cdot \mid x, y, u)$ with respect to a finite measure μ_Y on $\mathcal{Y}$, then

$$\big[\Phi_t(y, \xi, u, y')\big](\{x^k\}) = \frac{\sum_{i=1}^n q_t(x^k, y' \mid x^i, y, u)\, \xi^i}{\sum_{\ell=1}^n \sum_{i=1}^n q_t(x^\ell, y' \mid x^i, y, u)\, \xi^i}$$

Further specialization for hidden Markov models is possible, where

- The unobserved process $\{X_t\}_{t=1,\dots,T}$ is a Markov process itself
- For all $t = 1, \dots, T$, the observable part Y_t is a noisy function of X_t

Extended state history (including belief states):

$$h_t = (y_1, \xi_1, y_2, \xi_2, \dots, y_t, \xi_t) \in \mathbb{H}_t$$

Policies $\pi = (\pi_1, \dots, \pi_T)$ with decision rules $\pi_t(h_t) \in U_t(y_t)$

Markov Policy

For all $h_t, h'_t \in \mathbb{H}_t$, if $y_t = y'_t$ and $\xi_t = \xi'_t$, then $\pi_t(h_t) = \pi_t(h'_t) = \pi_t(y_t, \xi_t)$

Denote $v_t^\pi(h_t) = \rho_{t,T}^\pi\big(c_t(Y_t, \pi_t(H_t)), \dots, c_T(Y_T, \pi_T(H_T))\big)(h_t)$.

A family of dynamic risk measures $\{\rho_{t,T}^\pi\}_{t=1,\dots,T}^{\pi\in\Pi}$ is Markovian if for any Markov policy $\pi \in \Pi$, all $t = 1, \dots, T$, and all $h_t, h'_t \in \mathbb{H}_t$ such that $y_t = y'_t$ and $\xi_t = \xi'_t$, we have

$$v_t^\pi(h_t) = v_t^\pi(h'_t)$$

Thus for a Markovian policy Π, we can write $v_t^\pi(y_t, \xi_t)$ instead of $v_t^\pi(h_t)$

We denote the conditional distributions of Y_{t+1}, given the observations up to time t, by $\Upsilon_t(g_t, u_t)$.

If a family of dynamic risk measures $\left\{\rho^{\pi}_{t,T}\right\}^{\pi\in\Pi}_{t=1,\dots,T}$ is translation-invariant, stochastically conditionally time-consistent and Markov, then functionals $\sigma_t : \left\{(y, \xi, \Upsilon_t(y, \xi, u)) \;\middle|\; y \in \mathcal{Y}, \xi \in \mathcal{P}(\mathcal{X}), u \in \mathcal{U}_t(y)\right\} \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R}$ exist such that for all $t = 1, \dots, T-1$, for all $(y_t, \xi_t) \in \mathcal{Y} \times \mathcal{P}(\mathcal{X})$, the functional $\sigma_t(y_t, \xi_t, \cdot, \cdot)$ satisfies the law invariance property, and for all π all $h_t \in \mathbb{H}_t$, we have

$$
\begin{aligned}
v^{\pi}_t(h_t) &= c_t(y_t, \pi_t(h_t)) \\
&\quad + \sigma_t\Big(y_t,\ \xi_t,\ \Upsilon_t(y_t, \xi_t, \pi_t(h_t)), y' \mapsto v^{\pi}_{t+1}\left(h_t, y', \Phi_t\big(y_t, \xi_t, \pi_t(h_t), y'\big)\right)\Big)
\end{aligned}
$$

The risk model can be formulated in terms of static risk measures on the space of functions of y, rather than functions of (y, ξ).

We denote the conditional distributions of Y_{t+1}, given the observations up to time t, by $\Upsilon_t(g_t, u_t)$.

If a family of dynamic risk measures $\{\rho^{\pi}_{t,T}\}^{\pi \in \Pi}_{t=1,\dots,T}$ is translation-invariant, stochastically conditionally time-consistent and Markov, then functionals $\sigma_t : \{(y, \xi, \Upsilon_t(y, \xi, u)) \mid y \in \mathcal{Y}, \xi \in \mathcal{P}(\mathcal{X}), u \in \mathcal{U}_t(y)\} \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R}$ exist such that for all $t = 1, \dots, T-1$, for all $(y_t, \xi_t) \in \mathcal{Y} \times \mathcal{P}(\mathcal{X})$, the functional $\sigma_t(y_t, \xi_t, \cdot, \cdot)$ satisfies the law invariance property, and for all π all $h_t \in \mathbb{H}_t$, we have (for Markovian policy π)

$$v^{\pi}_t(y_t, \xi_t) = c_t(y_t, \pi_t(y_t, \xi_t)) + \sigma_t\Big(y_t,\ \xi_t,\ \Upsilon_t(y_t, \xi_t, \pi_t(y_t, \xi_t)), y' \mapsto v^{\pi}_{t+1}\big(y', \Phi_t(y_t, \xi_t, \pi_t(y_t, \xi_t), y')\big)\Big)$$

The risk model can be formulated in terms of static risk measures on the space of functions of y, rather than functions of (y, ξ).

Dynamic Programming

Risk-averse optimal control problem:

$$\min_{\pi} \ \rho^{\pi}_{1,T}\big\{c_1(y_1,u_1), c_2(y_2,u_2), \dots, c_T(y_T,u_T)\big\}$$

Theorem

If the risk measure is Markovian (+ general conditions), then the optimal solution is given by the dynamic programming equations:

$$v_T^*(y,\xi) = \min_{u\in\mathcal{U}_T(x)} c_T(y,u), \quad y\in\mathcal{Y}, \quad \xi\in\mathcal{P}(\mathcal{X})$$

$$v_t^*(y,\xi) = \min_{u\in\mathcal{U}_t(y)} \Big\{c_t(y,u) + \sigma_t\Big(y,\xi,\int_{\mathcal{Y}} K_t^X(x,\cdot,u)\ d\xi, y'\mapsto v_{t+1}^*\big(y',\Phi_t(x,\xi,u,y')\big)\Big)\Big\},$$

$$y\in\mathcal{Y}, \quad \xi\in\mathcal{P}(\mathcal{X}), \quad t=T-1,\dots,1$$

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

Risk-Averse Clinical Trials (with D. Dentcheva and Curtis McGinity)

- In stages $t = 1, \dots, T$ of the trials successive patients are given drugs (cytotoxic agents), to which severe toxic response (death) is possible
- Probability of toxic response depends on the unknown optimal dose η^* and the administered dose (control) u:

$$F(u,\eta) = \frac{1}{1 - e^{-\varphi(u,\eta)}}$$

- The "belief state" s_t, the probability distribution of the unknown optimal dose, is the current state of the system
- The state evolves according to Bayes operator, depending on the response of the patient: for $\eta \in \mathcal{X}$ (the range of doses)

$$s_{t+1}(\eta) \sim \begin{cases} F(u_t,\eta)\, s_t(\eta) & \text{if toxic} \\ \big(1 - F(u_t,\eta)\big)\, s_t(\eta) & \text{if not toxic} \end{cases}$$

- Cost per stage: $c_t(s_t, u_t) = \gamma_t \mathbb{E}_{s_t}\big[|u_t - \eta|\big]$ (other forms possible)

Medical ethics naturally motivates risk-averse control

Total Cost Models

Find the best policy $\Pi = (\pi_1, \ldots, \pi_T)$ to determine doses $u_t = \pi_t(s_t)$

Expected Value Model

$$\min_{\Pi} \mathbb{E}\left[\sum_{t=1}^{T+1} \gamma_t |u_t - \eta^*|\right]$$

γ_{T+1} is the weight of the final recommendation u_{T+1}

Risk-Averse Value Model

$$\min_{\Pi} \varrho_{1,T+1}\left[\sum_{t=1}^{T+1} \gamma_t |u_t - \eta^*|\right]$$

Two sources of risk

- Unknown state η^* (only distribution s_t available)
- Unknown evolution due to random response r_t

Limited Lookahead Policies

At each time t, assume that this is the last test before the final recommendation, and solve the two-stage problem

Risk-Neutral

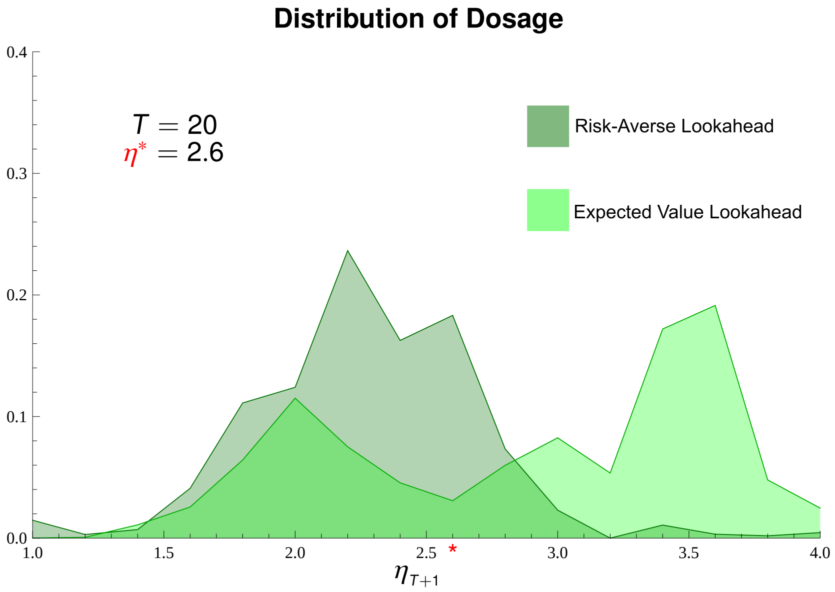
$$\min_{u_t} \mathbb{E}_{s_t} \left\{ \gamma_t |u_t - \eta| + \overline{\gamma}_{t+1} \; \mathbb{E}_{r_t} \left[\min_{u_{t+1}} \mathbb{E}_{s_{t+1}} |u_{t+1} - \eta| \right] \right\}$$

Risk-Averse

$$\min_{u_t} \rho_{s_t} \left\{ \gamma_t |u_t - \eta| + \overline{\gamma}_{t+1} \; \rho_{r_t} \left[\min_{u_{t+1}} \mathbb{E}_{s_{t+1}} |u_{t+1} - \eta| \right] \right\}$$

$$\overline{\gamma}_{t+1} = \sum_{\tau=t+1}^{T+1} \gamma_\tau \qquad \text{(weight of the future)}$$

Simulation Results for Expected Value and Risk-Averse Policies



References

- A. Ruszczyński, Risk-averse dynamic programming for Markov decision processes, *Mathematical Programming*, Series B 125 (2010) 235–261
- Ö. Çavuş and A. Ruszczyński, Computational methods for risk-averse undiscounted transient Markov models, *Operations Research*, 62 (2), 2014, 401–417.
- Ö. Çavuş and A. Ruszczyński, Risk-averse control of undiscounted transient Markov models, *SIAM J. on Control and Optimization*, 52(6), 2014, 3935–3966.
- J. Fan and A. Ruszczyński, Process-based risk measures for observable and partially observable discrete-time controlled systems, *submitted for publication*.
- C. McGinity, D. Dentcheva and A. Ruszczyński, Risk-averse approximate dynamic programming for partially observable systems with application to clinical trials, *in preparation*.