Risk-Averse Control
of Partially Observable Markov Systems

Andrzej Ruszczyński

RUTGERS

Workshop on Dynamic Multivariate Programming
Vienna, March 2018

Partially Observable Discrete-Time Models

- Markov Process: $\{X_t, Y_t\}_{t=1,\dots,T}$ on the Borel state space $\mathcal{X} \times \mathcal{Y}$
- The process $\{X_t\}$ is observable, while $\{Y_t\}$ is not observable
- Control sets: $U_t : \mathcal{X} \rightrightarrows \mathcal{U}$, $t = 1, \dots, T$
- Transition kernel: $\mathbb{P}[(X_{t+1}, Y_{t+1}) \in C \mid x_t, y_t, u_t] = Q_t(x_t, y_t, u_t)(C)$
- Costs: $Z_t = c_t(X_t, Y_t, U_t)$, $t = 1, \dots, T$

Two relevant filtrations

- $\{\mathcal{F}_t^{X,Y}\}$ defined by the full state process $\{X_t, Y_t\}$
- $\{\mathcal{F}_t^{X}\}$ defined by the observed process $\{X_t\}$

Space of costs: $\mathcal{Z}_t = \Big\{ Z : \Omega \to \mathbb{R} \,\Big|\, Z \text{ is } \mathcal{F}_t^{X,Y}\text{-measurable and bounded} \Big\}$

Classical Problem:

$$\min \ \mathbb{E}\left\{ c_1(X_1, Y_1, U_1) + c_2(X_2, Y_2, U_2) + \cdots + c_T(X_T, Y_T, U_T) \right\}$$

- Markov Process: $\{X_t, Y_t\}_{t=1,\dots,T}$ on the Borel state space $\mathcal{X} \times \mathcal{Y}$
- The process $\{X_t\}$ is observable, while $\{Y_t\}$ is not observable
- Control sets: $U_t : \mathcal{X} \rightrightarrows \mathcal{U}$, $t = 1, \dots, T$
- Transition kernel: $\mathbb{P}[(X_{t+1}, Y_{t+1}) \in C \mid x_t, y_t, u_t] = Q_t(x_t, y_t, u_t)(C)$
- Costs: $Z_t = c_t(X_t, Y_t, U_t)$, $t = 1, \dots, T$

Two relevant filtrations

- $\{\mathcal{F}_t^{X,Y}\}$ defined by the full state process $\{X_t, Y_t\}$
- $\{\mathcal{F}_t^{X}\}$ defined by the observed process $\{X_t\}$

Space of costs: $\mathcal{Z}_t = \Big\{ Z : \Omega \to \mathbb{R} \,\Big|\, Z \text{ is } \mathcal{F}_t^{X,Y}\text{-measurable and bounded} \Big\}$

Risk-Averse Problem:

$$\min \ \rho_{1,T}\big\{c_1(X_1, Y_1, U_1), c_2(X_2, Y_2, U_2), \dots, c_T(X_T, Y_T, U_T)\big\}$$

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: $\mathcal{Z}_t$ of $\mathcal{F}_t$-measurable functions and $\mathcal{Z}_{t,T} = \mathcal{Z}_t \times \cdots \times \mathcal{Z}_T$

Dynamic Risk Measure

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

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

Local property: For all $A \in \mathcal{F}_t$

$$\rho_{t,T}(\mathbb{1}_A Z) = \mathbb{1}_A \rho_{t,T}(Z)$$

Time Consistency and Nested Representation

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

$$Z_t = W_t \quad \text{and} \quad \rho_{t+1,T}(Z_{t+1}, \dots, Z_T) \le \rho_{t+1,T}(W_{t+1}, \dots, W_T)$$

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

Define one-step mappings: $\quad \rho_t(Z_{t+1}) = \rho_{t,T}(0, Z_{t+1}, 0, \dots, 0)$

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

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

Issues with General Theory in the Markov Setting

- Probability measure P^{Π}, processes X_t^{Π} and Z_t^{Π} depend on policy Π
- We have to deal with a family of risk measures $\rho_{t,T}^{\Pi}(\cdot)$
- The values of the risk measures may depend on history, and Markov policies cannot be expected
- The cost may not be observable

Motivating Example

$\mathcal{X} = \{0,1\}$, $T = 2$, and $Z_t = Z_t(x_t)$ (cost depends on state).
Consider the risk measure

$$\rho_{2,2}(Z_2)(x_1, x_2) = Z_2(x_2)$$

$$\rho_{1,2}(Z_1, Z_2)(x_1) = Z_1(x_1) + Z_2(x_1) \qquad \text{(assumes that } x_1 \text{ will not change)}$$

It is time-consistent and has the normalization, translation, and local properties.
As $\rho_{1,2}$ does not depend on the distribution of x_2, given x_1, it is useless for controlling Markov models. In fact, it is much worse than expectation.

Conditional Risk Evaluators

Space of observable random variables:

$$\mathcal{S}_t = \left\{ S : \Omega \to \mathbb{R} \,\middle|\, S \text{ is } \mathcal{F}_t^X\text{-measurable and bounded} \right\}, \quad t = 1, \dots, T$$

A mapping $\rho_{t,T} : \mathcal{Z}_t \times \cdots \times \mathcal{Z}_T \to \mathcal{S}_t$ is a conditional risk evaluator

(i) It is monotonic if $Z_s \le W_s$ for all $s = t, \dots, T$, implies that

$$\rho_{t,T}(Z_t, \dots, Z_T) \le \rho_{t,T}(W_t, \dots, W_T)$$

(ii) It is normalized if $\rho_{t,T}(0, \dots, 0) = 0$;

(iii) It is translation equivariant if $\forall (Z_t, \dots, Z_T) \in \mathcal{S}_t \times \mathcal{Z}_{t+1} \times \cdots \times \mathcal{Z}_T$,
$\rho_{t,T}(Z_t, \dots, Z_T) = Z_t + \rho_{t,T}(0, Z_{t+1}, \dots, Z_T)$;

(iv) It is decomposable if a mapping $\rho_t : \mathcal{Z}_t \to \mathcal{S}_t$ exists such that:

$$\rho_t(Z_t) = Z_t, \quad \forall Z_t \in \mathcal{S}_t,$$
$$\rho_{t,T}(Z_t, \dots, Z_T) = \rho_t(Z_t) + \rho_{t,T}(0, Z_{t+1}, \dots, Z_T), \quad \forall\, Z \in \mathcal{Z}_{t,T}$$

Risk Filters and their Time Consistency

A risk filter $\{\rho_{t,T}\}_{t=1,\dots,T}$ is a sequence of conditional risk evaluators $\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{S}_t$.

We have index risk filters by policy π, because π affects the measure P^π

History: $H_t^\pi = (X_1, X_2^\pi, \dots, X_t^\pi)$, $h_t = (x_1, x_2, \dots, x_t)$

A family of risk filters $\{\rho_{t,T}^\pi\}_{t=1,\dots,T}^{\pi\in\Pi}$ is stochastically conditionally time consistent if for any $\pi, \pi' \in \Pi$, for any $1 \le t < T$, for all $h_t \in \mathcal{X}^t$, all $(Z_t, \dots, Z_T) \in \mathcal{Z}_{t,T}$ and all $(W_t, \dots, W_T) \in \mathcal{Z}_{t,T}$, the conditions

$$Z_t = W_t$$

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

imply

$$\rho_{t,T}^\pi(Z_t, Z_{t+1}, \dots, Z_T)(h_t) \le \rho_{t,T}^{\pi'}(W_t, W_{t+1}, \dots, W_T)(h_t)$$

The relation $\preceq_{\text{st}}$ is the conditional stochastic order

Bayes Operator

Belief State: Conditional distribution of Y_t given initial distribution ξ_1 and history $g_t = (\xi_1, x_1, u_1, x_2, \ldots, u_{t-1}, x_t)$

$$\big[\varXi_t(g_t)\big](A) = \mathbb{P}[Y_t \in A \mid g_t], \quad \forall\, A \in \mathcal{B}(\mathcal{Y}), \quad t = 1, \ldots, T$$

Conditional distribution of the observable part:

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

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

Transition of the belief state - Bayes operator

$$\varXi_{t+1}(g_{t+1}) = \varPhi_t(x_t, \varXi_t(g_t), u_t, x_{t+1})$$

Example: $\mathcal{Y} = \{y^1, \ldots, y^n\}$ and $Q_t(x, y, u)$ has density $q_t(x', y' | x, y, u)$

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

Extended state history (including belief states):

$$h_t = (x_1, \xi_1, x_2, \xi_2, \ldots, x_t, \xi_t) \in \mathbb{H}_t$$

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

Markov Policy

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

Policy value function:

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

A family of risk filters $\left\{\rho_{t,T}^\pi\right\}_{t=1,\ldots,T}^{\pi\in\Pi}$ is Markov if for all Markov policies $\pi \in \Pi$, for all $h_t = (x_1, \ldots, x_t)$ and $h'_t = (x'_1, \ldots, x'_t)$ in $\mathcal{X}^t$ such that $x_t = x'_t$ and $\xi_t = \xi'_t$, we have

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

A family of risk filters $\left\{\rho_{t,T}^{\pi}\right\}_{t=1,\dots,T}^{\pi\in\Pi}$ is normalized, translation-invariant, stochastically conditionally time consistent, decomposable, and Markov if and only if transition risk mappings exist:

$$\sigma_t : \left\{\left(x_t, \xi_t, Q_t^{\pi}(h_t)\right) : \pi \in \Pi,\ h_t \in \mathcal{X}^t\right\} \times \mathcal{V} \to \mathbb{R}, \quad t = 1 \dots T-1,$$

(i) $\sigma_t(x, \xi, \cdot, \cdot)$ is normalized and strongly monotonic with respect to stochastic dominance

(ii) for all $\pi \in \Pi$, for all $t = 1, \dots, T-1$, and for all $h_t \in \mathcal{X}^t$,

$$v_t^{\pi}(h_t) = r_t(x_t, \xi_t, \pi_t(h_t)) + \sigma_t\left(x_t, \xi_t, Q_t^{\pi}(h_t), v_{t+1}^{\pi}(h_t, \cdot)\right)$$

Evaluation of a Markov policy π:

$$v_t^{\pi}(x_t, \xi_t) = r_t(x_t, \xi_t, \pi_t(x_t, \xi_t)) + \sigma_t\left(x_t, \xi_t, Q_t^{\pi}(x_t, \xi_t), x' \mapsto v_{t+1}^{\pi}(x', \Phi_t(x_t, \xi_t, \pi_t(x_t, \xi_t), x'))\right)$$

Examples of Transition Risk Mappings

Average Value at Risk

$$\sigma(x,\xi,m,v) = \min_{\eta\in\mathbb{R}} \left\{ \eta + \frac{1}{\alpha(x,\xi)} \int_{\mathcal{X}} \big(v(x') - \eta\big)_+ \; m(dx') \right\}$$

where $\alpha(x,\xi) \in [\alpha_{\min}, \alpha_{\max}] \subset (0,1)$.

Mean–Semideviation of Order p

$$\sigma(x,\xi,m,v) = \underbrace{\int_{\mathcal{X}} v(x') \; m(dx')}_{\mathbb{E}_m[v]} + \kappa(x,\xi) \left(\int_{\mathcal{X}} \big(v(x') - \mathbb{E}_m[v]\big)_+^p \; m(dx') \right)^{\frac{1}{p}}$$

where $\kappa(x,\xi) \in [0,1]$.

Entropic Mapping

$$\sigma(x,\xi,m,v) = \frac{1}{\gamma(x,\xi)} \ln\left(\mathbb{E}_m\left[e^{\gamma(x,\xi)\, v(x')} \right] \right), \quad \gamma(x,\xi) > 0$$

Risk-averse optimal control problem:

$$\min_{\pi} \ \rho^{\pi}_{1,T}\big\{c_1(X_1,Y_1,U_1), c_2(X_2,Y_2,U_2),\dots,c_T(X_T,Y_T,U_T)\big\}$$

Theorem

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

$$v^*_T(x,\xi) = \min_{u\in\mathcal{U}_T(x)} r_T(x,\xi,u), \quad x\in\mathcal{X}, \quad \xi\in\mathcal{P}(\mathcal{X})$$

$$\begin{aligned} v^*_t(x,\xi) = \min_{u\in\mathcal{U}_t(x)} \Big\{ & r_t(x,\xi,u) + \\ & \sigma_t\Big(x,\xi,\int_{\mathcal{Y}} K^X_t(x,y,u)\,\xi(dy), x'\mapsto v^*_{t+1}\big(x',\Phi_t(x,\xi,u,x')\big)\Big)\Big\}, \\ & x\in\mathcal{X}, \quad \xi\in\mathcal{P}(\mathcal{Y}), \quad t=T-1,\dots,1 \end{aligned}$$

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

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

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

- The "belief state" ξ_t, the conditional 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 $\eta \in \mathcal{Y}$ (the range of doses)

$$\xi_{t+1}(\eta) \sim \begin{cases} F(u_t, \eta)\, \xi_t(\eta) & \text{if toxic } \ (x_{t+1} = 1) \\ \big(1 - F(u_t, \eta)\big)\, \xi_t(\eta) & \text{if not toxic } \ (x_{t+1} = 0) \end{cases}$$

- Cost per stage: $c_t(\eta, u_t) = \gamma_t |u_t - \eta|$ (other forms possible)

Medical ethics naturally motivates risk-averse control

Total Cost Models

Find the best policy $\pi = (\pi_1, \dots, \pi_T)$ to determine doses $u_t = \pi_t(\xi_t)$

Expected Value Model

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

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

Risk-Averse Model

$$\min_{\pi \in \Pi} \rho^{\pi}_{1,T+1} \left[\left\{ \gamma_t |u_t - \eta^*| \right\}_{t=1,\dots,T+1} \right]$$

Two sources of risk

- Unknown state η^* (only belief state ξ_t available at time t)
- Unknown evolution of $\{\xi_t\}$ due to random responses of patients

Dynamic Programming Equations

- All memory is carried by the belief state ξ_t
- For each ξ_t and u_t, only two next states are possible, corresponding to $x_{t+1} = 0$ or 1

Simplified equation

$$v_t(\xi) = \min_u \left\{ r_t(\xi, u) + \sigma\left(\xi, \int_{\mathcal{Y}} \mathbb{P}[x' = 1 | y, u]\, \xi(dy), v^*_{t+1}\big(\Phi_t(x, \xi, u, \cdot)\big)\right) \right\}$$

Examples:

$$r_t(\xi, u) = \mathbb{E}_\xi\big[|u - \eta|\big]$$

$$\sigma\big(\xi, p, \varphi(\cdot)\big) = \mathbb{E}_\xi\Big[\max_{x' \in \{0,1\}} \varphi(x')\Big]$$

Any law invariant risk measure on the space of functions on U (for r_t) or on $U \times \{0, 1\}$ (in the case of σ_t) can be used here.

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

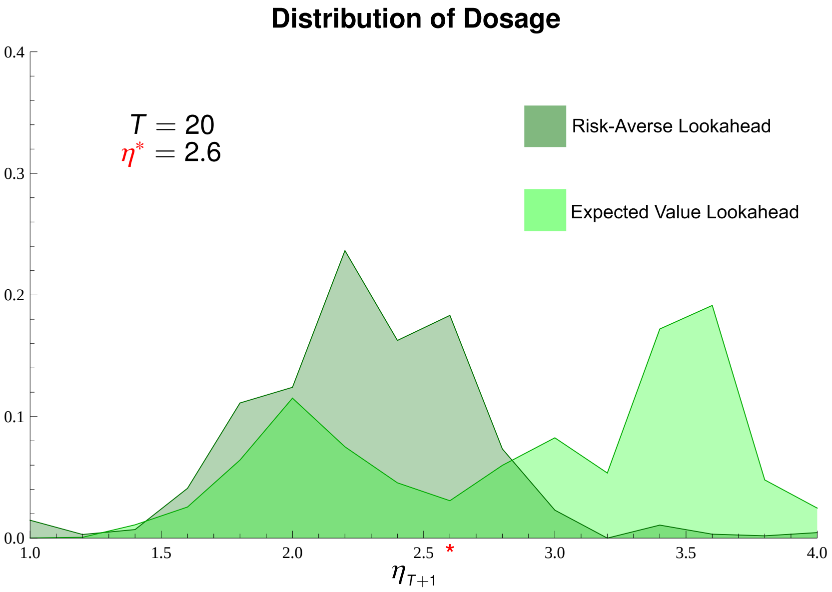
$$\min_{u_t} \mathbb{E}_{\xi_t} \left\{ \gamma_t |u_t - \eta| + \overline{\gamma}_{t+1} \; \mathbb{E}_{\text{response}} \left[\min_{u_{t+1}} \mathbb{E}_{\xi_{t+1}} |u_{t+1} - \eta| \right] \right\}$$

Risk-Averse

$$\min_{u_t} \mathbb{E}_{\xi_t} \left\{ \gamma_t |u_t - \eta| + \overline{\gamma}_{t+1} \; \max_{\text{response}} \left[\min_{u_{t+1}} \mathbb{E}_{\xi_{t+1}} |u_{t+1} - \eta| \right] \right\}$$

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

Simulation Results for Expected Value and Risk-Averse Policies



Machine Deterioration (with Jingnan Fan)

We consider the problem of minimizing costs of a machine in T periods.

Unobserved state: $y_t \in \{1, 2\}$, with 1 being the "good" and 2 the "bad" state
Observed state: x_t - cost incurred in the previous period
Control: $u_t \in \{0, 1\}$, with 0 meaning "continue", and 1 meaning "replace"

The dynamics of Y is Markovian, with the transition matrices $K^{[u]}$:

$$K^{[0]} = \begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix} \quad K^{[1]} = \begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix}$$

Distribution of costs:

$$\mathbb{P}\big[x_{t+1} \le C \,\big|\, y_t = i, u_t = 0\big] = \int_{-\infty}^{C} f_i(x)\, dx, \quad i = 1, 2$$

$$\mathbb{P}\big[x_{t+1} \le C \,\big|\, y_t = i, u_t = 1\big] = \int_{-\infty}^{C} f_1(x)\, dx, \quad i = 1, 2$$

Value and Policy Monotonicity

Belief state: $\xi_i \in [0,1]$ - conditional probability of the "good" state
The optimal value functions: $v_t^*(x,\xi) = x + w_t^*(\xi),\ t = 1,\dots,T+1$

Dynamic programming equations

$$w_t^*(\xi) = \min\Big\{R + \sigma\big(f_1, x' \mapsto x' + w_{t+1}^*(1-p)\big);$$
$$\sigma\big(\xi f_1 + (1-\xi)f_2, x' \mapsto x' + w_{t+1}^*(\Phi(\xi,x'))\big)\Big\},$$

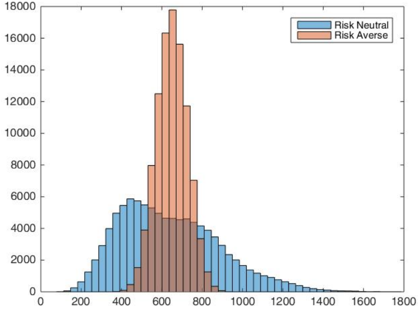
with the final stage value $w_{T+1}^*(\cdot) = 0$.

If $\frac{f_1}{f_2}$ is non-increasing, then the functions $w_t^*(\cdot)$ are non-increasing and thresholds $\xi_t^* \in [0,1],\ t = 1,\dots,T$ exist, such that the policy

$$u_t^* = \begin{cases} 0 & \text{if } \xi_t > \xi_t^*, \\ 1 & \text{if } \xi_t \le \xi_t^*, \end{cases}$$

is optimal

Cost distributions f_1 and f_2: uniform with $\int_0^\eta f_1(x)\,dx \preceq \int_0^\eta f_2(x)\,dx$
Transition risk mapping: mean–semideviation



Empirical distribution of the total cost for the risk-neutral model (blue) and the risk-averse model (orange)

Partially Observable Jump Process

The unobserved process $\{Y_t\}_{0\le t\le T}$: Finite state Markov jump process on the space $\mathcal{Y} = \{1,\dots,n\}$ with the generator $\varLambda(t)$:

$$\lambda_{ij}(t) = \begin{cases} \lim\limits_{\varepsilon\downarrow 0} \dfrac{1}{\varepsilon}\mathbb{P}\big[Y_{t+\varepsilon}=j \big| Y_t = i\big] & \text{if } j \neq i \\ -\sum_{k\neq i} \lambda_{ik}(t) & \text{if } j = i \end{cases}$$

The observed process $\{X_t\}_{0\le t\le T}$: Diffusion following the SDE

$$dX_t = A(Y_t,t)\,dt + B(t)\,dW_t, \quad 0 \le t \le T,$$

with the initial value x_0, independent of Y_0. $\{W_t\}$ is a Wiener process.

Random final cost: $\phi(Y_T)$

The Belief State Equation (Wonham, 1964)

Filtration of observable events: $\{\mathcal{F}_t^X\}_{0\le t\le T}$

The belief state: $\xi_i(t) = P\big[Y_t = i \mid \mathcal{F}_t^X\big], \quad i = 1,\dots,n,$

Belief State Equation

$$d\xi_i(s) = (\Lambda^*\xi)_i(s)\,ds + \xi_i(s)\frac{A(i,s)-\bar{A}(s)}{B(s)}\,d\overline{W}_s, \quad \xi_i(0) = p_i,$$

where

$$(\Lambda^*\xi)_i(s) = \sum_{j=1}^{n}\lambda_{ji}(s)\,\xi_j(s), \qquad \bar{A}(s) = \sum_{j=1}^{n} A(j,s)\,\xi_j(s),$$

and $\{\overline{W}_t\}_{0\le t\le T}$ is a Wiener process given by the formula

$$\overline{W}_t = \int_0^t \frac{dX_s - \bar{A}(s)\,ds}{B(s)}.$$

Suppose $\pi(t) = p$ and we use a Markov risk filter $\{\varrho_{t,T}\}_{0\le t\le T}$.
The value function for the final cost case:

$$V(t,p) = \varrho_{t,T}\left[\phi\big(Y_T^{t,p}\big)\right]$$

Structure of $\varrho_{t,T}(\cdot)$ [using Coquet, Hu, Mémin, Peng (2002)]

If the filter is monotonic, normalized, time consistent, and has the local property (+ minor growth conditions) then a driver $g : [0,T]\times\mathbb{R}\times\mathbb{R}^n \to \mathbb{R}$ exists, such that $\varrho_{t,T}\big[\phi\big(Y_T^{t,p}\big)\big] = V_t$, where (V,Z) solve backward stochastic differential equation

$$-dV_s = g(s,V_s,Z_s)\,ds - Z_s\,d\overline{W}_s,\quad s\in[t,T],\quad V_T = \varrho_{T,T}\left[\phi\big(Y_T^{t,p}\big)\right]$$

Under additional condition of law invariance of $\rho_{T,T}[\cdot]$, we obtain the following system.

Forward SDE for the belief state: For $i = 1, \dots, n$ and $0 \le t \le s \le T$

$$d\xi_i(s) = (\Lambda^* \xi)_i(s)\; ds + \xi_i(s) \frac{A(i,s) - \bar{A}(s)}{B(s)}\; d\overline{W}_s, \quad \xi_i(t) = p_i,$$

Backward SDE for the risk measure: for $0 \le t \le s \le T$

$$-dV_s = g(s, V_s, Z_s)\, ds - Z_s\, d\overline{W}_s, \quad s \in [t, T],$$

$$V_T = r_T\big(\phi, \xi(T)\big)$$

The functional $r_T(\cdot,\cdot)$ is a law invariant risk measure.

Functional with Running Cost:

$$Z_T = \int_0^T c\big(\xi(t)\big)\,dt + \phi(Y_T)$$

The value function:

$$V(t,p) = \varrho_{t,T}\left[Z_T\right]$$

Forward SDE for the belief state: For $i = 1, \dots, n$ and $0 \le t \le s \le T$

$$d\xi_i(s) = (\Lambda^*\xi)_i(s)\,ds + \xi_i(s)\frac{A(i,s) - \bar{A}(s)}{B(s)}\,d\overline{W}_s, \quad \xi_i(t) = p_i,$$

Backward SDE for the risk measure: for $0 \le t \le s \le T$

$$-dV_s = \Big[c\big(\xi(s)\big) + g(s, V_s, Z_s)\Big]\,ds - Z_s\,d\overline{W}_s, \quad s \in [t, T],$$

$$V_T = r_T\big(\phi, \xi(T)\big)$$

Controlled transition rates: $\lambda_{ij}(t,\zeta)$, $\zeta \in U$, where U is a bounded set. The rates are uniformly bounded.

Piecewise-constant control: For $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$, we define

$$\mathcal{U}_i^N = \left\{ u \in \mathcal{U} \,\middle|\, u(t) = u(t_j),\ \forall\, t \in [t_j, t_{j+1}),\ \forall\, j = i, \ldots, N-1 \right\}$$

where $\mathcal{U}$ is the set of U-valued processes, adapted to $\{\mathcal{F}_t^{\xi}\}_{0 \le t \le T}$.

Value function for a fixed control:

$$V^u(t_j, p) = \rho_{t_j, t_{j+1}} \left[\int_{t_j}^{t_{j+1}} c(\xi^{t_j,p;u}(s), u_r)\, ds + V^u\Big(t_{j+1}, \xi^{t_j,p;u}(t_{j+1})\Big) \right]$$

$\xi^{t_j,p;u}(\cdot)$ is the belief process restarted at t_j from value p, while the system is controlled by $u(\cdot) = u(t_j)$ in the interval $[t_j, t_{j+1})$.

Dynamic Programming

Optimal value function: $\hat{V}(t_j, p) = \inf_{u \in \mathcal{U}_j^N} V^u(t_j, p)$

$$\hat{V}(t_j, p) = \inf_{\zeta \in U} \rho_{t_j, t_{j+1}} \left[\int_{t_j}^{t_{j+1}} c(\xi^{t_j,p;\zeta}(r), \zeta)\, dr + \hat{V}\Big(t_{j+1}, \xi^{t_j,p;\zeta}(t_{j+1})\Big) \right]$$

Each $\rho_{t_j,t_{j+1}}[\cdot]$ is given by a controlled FBSDE system on $[t_j, t_{j+1}]$

$$\begin{aligned} d\xi_i^{t_j,p;\zeta}(s) &= (\Lambda^*(\zeta)\,\xi)_i(s)\, ds + \xi_i^{t_j,p;\zeta}(s)\frac{A(i,s) - \bar{A}(s)}{B(s)}\, d\overline{W}_s, \\ \xi_i^{t_j,p;\zeta}(t_j) &= p_i, \\ -dV_s &= \Big[c\Big(\xi^{t_j,p;\zeta}(s), \zeta\Big) + g(s, V_s, Z_s)\Big]\, ds - Z_s\, d\overline{W}_s, \\ V_{t_{j+1}} &= \hat{V}\Big(t_{j+1}, \xi^{t_j,p;\zeta}(t_{j+1})\Big) \end{aligned}$$

Further research: Numerical methods for this FBSDE system

References

- A. Ruszczyński, Risk-averse dynamic programming for Markov decision processes, *Mathematical Programming*, Series B 125 (2010) 235–261
- Ö. Çavuș and A. Ruszczyński, Computational methods for risk-averse undiscounted transient Markov models, *Operations Research*, 62 (2), 2014, 401–417.
- Ö. Çavuș 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, Risk measurement and risk-averse control of partially observable discrete-time Markov systems, *Mathematical Methods of Operations Research*, 2018, 1–24.
- C. McGinity, D. Dentcheva and A. Ruszczyński, Risk-averse approximate dynamic programming for partially observable systems with application to clinical trials, *in preparation*
- R. Yan, J. Yao, A. Ruszczyński, Risk Filtering and Risk-Averse Control of Partially Observable Markov Jump Processes, *submitted for publication*.
- D. Dentcheva and A. Ruszczyński, Risk measures for continuous-time Markov chains, *submitted for publication*.