Risk Sensitive Stochastic Games

Mrinal K. Ghosh
(Coauthor: A. Basu)

Department of Mathematics
Indian Institute of Science

Workshop on Game Theory and Mechanism Design
11-15 January, 2016

Problem Description

A two-person stochastic game is determined by six objects: (X, U, V, r_1, r_2, Q), where

- $X = \{1, 2, \cdots\}$ is the state space,
- U, V are action spaces of player 1 and 2 respectively, assumed to be compact metric spaces,
- $r_i : X \times U \times V \to \mathbb{R}$, $i = 1, 2$ is the one-stage cost function for player i, assumed to be bounded and continuous,
- $Q : X \times U \times V \to \mathcal{P}(X)$, is the transition stochastic kernel, assumed to be continuous in (u, v) in the topology of weak convergence.

Evolution of the system and information

The game is played as follows: At each stage players observe the current state $x \in X$ and then players independently choose actions $u \in U, \ v \in V$. As a result two things happen

- player $i, \ i = 1, 2$, pays an immediate cost $r_i(x, u, v)$
- the system moves to a new state $x' \in X$ with probability $Q(x'|x, u, v)$.

The whole process then repeats from the new state x'.
The available information at time $t = 0, 1, 2, \cdots$, is given by the history

$$h_t = (x_0, (u_0, v_0), x_1, (u_1, v_1), \cdots, (u_{t-1}, v_{t-1}), x_t) \in H_t$$

where $H_0 = X, \ H_t = H_{t-1} \times U \times V \times X, \ H_\infty = (U \times V \times X)^\infty$.

Strategies

- A strategy for player 1 is a sequence

$$\mu = \{\mu_t : H_t \to \mathcal{P}(U)\}$$

  Let $\Pi_i =$ the set of all strategies of player i.
- A Markov strategy for player 1 is given by

$$\mu_t : \mathbb{N} \times X \to \mathcal{P}(U)$$

- A stationary strategy for player 1 is given by

$$\mu : X \to \mathcal{P}(U)$$

- We denote the set of all Markov strategies by $\mathcal{M}_i$ and the set of all stationary strategies by $\mathcal{S}_i$ for the ith player.
- Given an initial distribution π_0 and a pair of strategies (μ, ν), the corresponding state and action process $\{X_t\}, \{U_t\}, \{V_t\}$ are defined on the canonical sample space $(H_\infty, \mathcal{B}(H_\infty), P^{\mu,\nu}_{\pi_0})$ via the standard projections:

$$X_t(h_\infty) = x_t, U_t(h_\infty) = u_t, V_t(h_\infty) = v_t.$$

  When $\pi_0 = \delta_x$, we write $P^{\mu,\nu}_x$.

Cost Evaluation Criteria

Risk-sensitive discounted cost

Let $\alpha \in (0,1)$ be the discount factor and $\theta \in (0,\Theta)$ the risk-sensitive parameter. The risk-sensitive discounted cost is given by

$$\rho_i^{\mu,\nu}(x) \;:=\; \frac{1}{\theta} \ln E_x^{\mu,\nu}\left[e^{\theta \sum_{t=0}^{\infty} \alpha^t r_i(X_t,U_t,V_t)}\right], \tag{1}$$

Definition 1

A pair of strategies (μ^*,ν^*) is called a Nash equilibrium if

$$\rho_1^{\mu^*,\nu^*}(x) \;\le\; \rho_1^{\mu,\nu^*}(x) \text{ for all } \mu \in \Pi_1 \text{ and } x \in X$$

and

$$\rho_2^{\mu^*,\nu^*}(x) \;\le\; \rho_2^{\mu^*,\nu}(x) \text{ for all } \nu \in \Pi_2 \text{ and } x \in X$$

Cost Evaluation Criteria

Risk-sensitive average cost

$$\beta_i^{\mu,\nu}(x) \;:=\; \limsup_{T\to\infty} \frac{1}{\theta T} \ln E_x^{\mu,\nu}\left[e^{\theta \sum_{t=0}^{T-1} r_i(X_t,U_t,V_t)}\right]. \qquad (2)$$

Remark

When the parameter $\theta \to 0$, we obtain the risk-neutral cost criteria, viz

$$J_i^{\mu,\nu}(x) \;:=\; E_x^{\mu,\nu}\left[\sum_{t=0}^{\infty} \alpha^t r_i(X_t, U_t, V_t)\right],$$

which is the discounted cost.
The averse cost is given by

$$L_i^{\mu,\nu}(x) \;:=\; \limsup_{T\to\infty} \frac{1}{T} E_x^{\mu,\nu}\left[\sum_{t=0}^{T-1} r_i(X_t, U_t, V_t)\right].$$

Analysis of discounted cost criterion

Since logarithm is an increasing function, it suffices to consider the (risk-sensitive) exponential cost criterion. For player i, the exponential cost is given by

$$\mathcal{J}_i^{\mu,\nu}(\theta,(x,t)) \;:=\; E_{x,t}^{\mu,\nu}\left[e^{\theta \sum_{s=t}^{\infty} \alpha^{s-t} r_i(X_s,U_s,V_s)}\right].$$

Dynamic programming equations

Given $(\mu,\nu) \in \mathcal{M}_1 \times \mathcal{M}_2$, consider the following equations

$$\begin{cases} \phi_1(\theta,(x,t)) = \inf_{\xi \in \mathcal{P}(U)} \Big[\int_U \int_V e^{\theta r_1(x,u,v)} \sum_{y \in X} \phi_1(\theta\alpha,(y,t+1)) \\ \qquad Q(y|x,u,v)\xi(du)\nu_t(x)(dv) \Big] \\ \text{with } \lim_{\theta \to 0} \phi_1(\theta,(x,t)) = 1. \end{cases}$$

and

Analysis of discounted cost criterion

Dynamic programming equations

$$\begin{cases} \phi_2(\theta,(x,t)) = \inf_{\chi \in \mathcal{P}(V)} \Big[\int_U \int_V e^{\theta r_2(x,u,v)} \sum_{y \in X} \phi_2(\theta\alpha,(y,t+1)) \\ \qquad Q(y|x,u,v)\mu_t(x)(du)\chi(dv) \Big] \\ \text{with } \lim_{\theta \to 0} \phi_2(\theta,(x,t)) = 1. \end{cases}$$

Analysis of discounted cost criterion

Theorem 2

Given $(\mu,\nu) \in \mathcal{M}_1 \times \mathcal{M}_2$, there exist unique bounded solutions to the above equations such that

$$\begin{aligned} \hat{\phi}_1[\nu](\theta,(x,t)) = & \inf_{\tilde{\mu}} \mathcal{J}_1^{\tilde{\mu},\nu}(\theta,(x,t)) \\ \hat{\phi}_2[\mu](\theta,(x,t)) = & \inf_{\tilde{\nu}} \mathcal{J}_2^{\mu,\tilde{\nu}}(\theta,(x,t)) \end{aligned}$$

Moreover there exist measurable maps

$$(\hat{\mu}[\nu],\hat{\nu}[\mu]) : (0,\Theta) \times (X \times \mathbb{N}) \to \mathcal{P}(U) \times \mathcal{P}(V)$$

such that

$$\begin{cases} \inf_{\xi \in \mathcal{P}(U)} \Big[\int_U \int_V e^{\theta r_1(x,u,v)} \sum_{y \in X} \hat{\phi}_1[\nu](\theta\alpha,(y,t+1)) Q(y|x,u,v) \xi(du) \nu_t(x)(dv) \Big] \\ = \int_U \int_V e^{\theta r_1(x,u,v)} \sum_{y \in X} \hat{\phi}_1[\nu](\theta\alpha,(y,t+1)) Q(y|x,u,v) \hat{\mu}[\nu](\theta,(x,t))(du) \nu_t(x)(dv) \end{cases} \tag{3}$$

and

Analysis of discounted cost criterion

Theorem 2 Continued

$$\begin{cases} \inf_{\chi \in \mathcal{P}(V)} \left[\int_U \int_V e^{\theta r_2(x,u,v)} \sum_{y \in X} \hat{\phi}_2[\mu](\theta\alpha, (y, t+1)) Q(y|x,u,v) \mu_t(x)(du) \chi(dv) \right] \\ = \int_U \int_V e^{\theta r_2(x,u,v)} \sum_{y \in X} \hat{\phi}_2[\mu](\theta\alpha, (y, t+1)) Q(y|x,u,v) \mu_t(x)(du) \hat{\nu}[\mu](\theta, (x,t))(dv). \end{cases} \tag{4}$$

Hence given $(\mu, \nu) \in \mathcal{M}_1 \times \mathcal{M}_2$ and $\theta \in (0, \Theta)$, the minimizing strategies $\{\mu_t^*[\nu]\} \in \mathcal{M}_1$, $\{\nu_t^*[\mu]\} \in \mathcal{M}_2$ are given by

$$\begin{aligned} \mu_t^*[\nu] &= \hat{\mu}[\nu](\theta\alpha^t, (X_t, t)) \\ \nu_t^*[\mu] &= \hat{\nu}[\mu](\theta\alpha^t, (X_t, t)). \end{aligned}$$

Thus $\mu_t^*[\nu]$ (resp. $\nu_t^*[\mu]$) is an optimal response (resp. player 2) corresponding to $\nu \in \mathcal{M}_2$ (resp. $\mu \in \mathcal{M}_1$).

Analysis of discounted cost criterion

Next define

$$H_i : \mathcal{M}_j \to 2^{\mathcal{M}_i}, \; i = 1, 2, \; i \neq j$$

by

$$H_1[\nu] = \{\mu_t^*[\nu] \in \mathcal{M}_1 : \mu_t^*[\nu] \text{ satisfies } (3)\}$$

$$H_2[\mu] = \{\nu_t^*[\mu] \in \mathcal{M}_2 : \nu_t^*[\mu] \text{ satisfies } (4)\}$$

Let $H = H_1 \times H_2 : \mathcal{M}_1 \times \mathcal{M}_2 \to 2^{\mathcal{M}_1 \times \mathcal{M}_2}$ be given by

$$H(\mu, \nu) = H_1[\nu] \times H_2[\mu]$$

Theorem 3

Given $\theta \in (0, \Theta)$, there exists a Nash equilibrium in $\mathcal{M}_1 \times \mathcal{M}_2$.

Proof.

Follows by applying a standard fixed point theorem. □

Analysis of discounted cost criterion

Remark

Comparison with risk-neutral discounted case. In this case the dynamic programming equations are as follows: for $(\mu, \nu) \in \mathcal{S}_1 \times \mathcal{S}_2$, consider

$$\psi_1[\nu](x) = \inf_{\tilde{\mu} \in \Pi_1} J_1^{\tilde{\mu},\nu}(x)$$
$$\psi_2[\mu](x) = \inf_{\tilde{\nu} \in \Pi_2} J_2^{\mu,\tilde{\nu}}(x).$$

Then $\psi_1[\nu](x)$ is the unique bounded solution of

$$\begin{cases} \psi_1[\nu](x) = \inf_{\mu \in \mathcal{P}(U)} \left[\int_U \int_V \left\{ r_1(x,u,v) + \alpha \sum_{y \in X} \psi_1[\nu](y) Q(y|x,u,v) \right\} \mu(du)\nu(x)(dv) \right] \\ \qquad = \int_U \int_V \left\{ r_1(x,u,v) + \alpha \sum_{y \in X} \psi_1[\nu](y) Q(y|x,u,v) \right\} \mu^*[\nu](du)\nu(x)(dv), \end{cases}$$

and

Analysis of discounted cost criterion

Remark continued

$\psi_2[\mu](x)$ is the unique bounded solution of

$$\begin{cases} \psi_2[\mu](x) = \inf_{\nu \in \mathcal{P}(V)} \Big[\int_U \int_V \Big\{ r_2(x,u,v) + \alpha \sum_{y \in X} \psi_2[\mu](y) Q(y|x,u,v) \Big\} \mu(x)(du) \nu(dv) \Big] \\ \qquad = \int_U \int_V \Big\{ r_2(x,u,v) + \alpha \sum_{y \in X} \psi_2[\mu](y) Q(y|x,u,v) \Big\} \mu(x)(du) \nu^*[\mu](dv), \end{cases}$$

Furthermore $\mu^* \in \mathcal{S}_1$ (resp. $\nu^* \in \mathcal{S}_2$) is an optimal response of player 1 (resp. ν^* of player 2) given player 2 (resp. player 1) is employing $\nu^* \in \mathcal{S}_2$ (resp. $\mu^* \in \mathcal{S}_1$).
Using this one can show the existence of a Nash equilibrium in stationary strategies.

Assumption

(i) The process $\{X_t\}$ is an irreducible, aperiodic Markov chain under any pair of stationary Markov strategies.
(ii) (*Lyapunov stability*): There exist constants $\eta < 1,\ b < \infty$ and a function $V : X \to [1, \infty)$ such that

$$\sum_{y \in X} V(y)Q(y|x, u, v) \quad \leq \quad \eta V(x) + bI_C(x).$$

Let

$$B_V(X) = \left\{ f : X \to \mathbb{R} |\ \sup_x \frac{|f(x)|}{V(x)} < \infty \right\}$$

Risk-sensitive average cost

Dynamic programming equations

Given strategies $(\mu, \nu) \in \mathcal{S}_1 \times \mathcal{S}_2$, consider the following equations

$$\begin{cases} e^{\theta\lambda_1 + V_1(\theta,x)} = \inf_{\xi \in \mathcal{P}(U)} \left[\int_U \int_V e^{\theta r_1(x,u,v)} \sum_{y \in X} e^{V_1(\theta,y)} Q(y|x,u,v) \xi(du) \nu(x)(dv) \right] \\ \quad = \int_U \int_V e^{\theta r_1(x,u,v)} \sum_{y \in X} e^{V_1(\theta,y)} Q(y|x,u,v) \mu^*[\nu](x)(du) \nu(x)(dv), \text{ say} \end{cases}$$

and

$$\begin{cases} e^{\theta\lambda_2 + V_2(\theta,x)} = \inf_{\chi \in \mathcal{P}(V)} \left[\int_U \int_V e^{\theta r_2(x,u,v)} \sum_{y \in X} e^{V_2(\theta,y)} Q(y|x,u,v) \mu(x)(du) \chi(dv) \right] \\ \quad = \int_U \int_V e^{\theta r_2(x,u,v)} \sum_{y \in X} e^{V_2(\theta,y)} Q(y|x,u,v) \mu(x)(du) \nu^*[\mu](x)(dv), \text{ say.} \end{cases}$$

Risk-sensitive averse cost

- Then $\lambda_1 = \lambda_1[\nu]$ is the optimal (risk-sensitive) average cost for player 1 if player 2 employs ν and $\mu^*[\nu] \in \mathcal{S}_1$ is an optimal response of player 1.
- Similarly, $\lambda_2 = \lambda_2[\mu]$ is the optimal average cost for player 2 if player 1 employs μ and $\nu^*[\mu] \in \mathcal{S}_2$ is an optimal response of player 2.
- Using the above, we have the following theorem:

Risk-sensitive average cost

Theorem 4

There exist scalars λ_1^*, λ_2^* strategies $(\mu^*, \nu^*) \in \mathcal{S}_1 \times \mathcal{S}_2$ and functions $V_1^*(\theta, \cdot)$, $V_2^*(\theta, \cdot) \in B_V(X)$ such that

$$\begin{cases} e^{\theta\lambda_1^* + V_1^*(\theta,x)} = \inf_{\xi \in \mathcal{P}(U)} \left[\int_U \int_V e^{\theta r_1(x,u,v)} \sum_{y \in X} e^{V_1^*(\theta,y)} Q(y|x,u,v) \xi(du) \nu^*(x)(dv) \right] \\ \qquad = \int_U \int_V e^{\theta r_1(x,u,v)} \sum_{y \in X} e^{V_1^*(\theta,y)} Q(y|x,u,v) \mu^*(x)(du) \nu^*(x)(dv) \end{cases}$$

and

$$\begin{cases} e^{\theta\lambda_2^* + V_2^*(\theta,x)} = \inf_{\chi \in \mathcal{P}(V)} \left[\int_U \int_V e^{\theta r_2(x,u,v)} \sum_{y \in X} e^{V_2^*(\theta,y)} Q(y|x,u,v) \mu^*(x)(du) \chi(dv) \right] \\ \qquad = \int_U \int_V e^{\theta r_2(x,u,v)} \sum_{y \in X} e^{V_2(\theta,y)} Q(y|x,u,v) \mu^*(x)(du) \nu^*(x)(dv). \end{cases}$$

Moreover, $(\mu^*, \nu^*) \in \mathcal{S}_1 \times \mathcal{S}_2$ is a Nash equilibrium and $(\lambda_1^*, \lambda_2^*)$ corresponding Nash Values.

Zero-Sum Case

- The usual zero sum game mean

$$r_1(x,u,v) + r_2(x,u,v) = 0$$

Thus

$$r_1(x,u,v) = -r_2(x,u,v) := r(x,u,v)$$

- In this case player 1 is risk-averse whereas player 2 is risk-seeking. This case again leads to coupled dynamic programming equations as in the non-zero sum case
- Suppose player 1 minimizes

$$\limsup_{T\to\infty} \frac{1}{\theta T} \ln E_x^{\mu,\nu}\left[e^{\theta \sum_{t=0}^{T-1} r(X_t,U_t,V_t)} \right],$$

over his strategies and player 2 tries to maximize the same.

Zero-Sum Case

- Then one gets a value of this game and saddle point strategies via the following Shapley equations:

$$\begin{cases} e^{\theta\lambda+V(\theta,x)} = \inf_{\xi\in\mathcal{P}(U)} \sup_{\chi\in\mathcal{P}(V)} \left[\int_U \int_V e^{\theta r(x,u,v)} \sum_{y\in X} e^{V(\theta,y)} Q(y|x,u,v)\xi(du)\chi(dv) \right] \\ \qquad\qquad\quad = \sup_{\chi\in\mathcal{P}(V)} \inf_{\xi\in\mathcal{P}(U)} \left[\int_U \int_V e^{\theta r(x,u,v)} \sum_{y\in X} e^{V(\theta,y)} Q(y|x,u,v)\xi(du)\chi(dv) \right] \end{cases}$$

- If the above equation has a suitable solution $(\lambda, V(\theta,x))$ then λ is the value of the game for the average cost.
- Furthermore if $(\mu^*,\nu^*) \in \mathcal{S}_1 \times \mathcal{S}_2$ be such that

Zero-Sum Case

- Furthermore if $(\mu^*, \nu^*) \in \mathcal{S}_1 \times \mathcal{S}_2$ be such that

$$\begin{cases} \inf_{\xi \in \mathcal{P}(U)} \sup_{\chi \in \mathcal{P}(V)} \left[\int_U \int_V e^{\theta r(x,u,v)} \sum_{y \in X} e^{V(\theta,y)} Q(y|x,u,v) \xi(du) \chi(dv) \right] \\ = \sup_{\chi \in \mathcal{P}(V)} \left[\int_U \int_V e^{\theta r(x,u,v)} \sum_{y \in X} e^{V(\theta,y)} Q(y|x,u,v) \mu^*(x)(du) \chi(dv) \right] \end{cases}$$

and

$$\begin{cases} \sup_{\chi \in \mathcal{P}(V)} \inf_{\xi \in \mathcal{P}(U)} \left[\int_U \int_V e^{\theta r(x,u,v)} \sum_{y \in X} e^{V(\theta,y)} Q(y|x,u,v) \xi(du) \chi(dv) \right] \\ = \inf_{\xi \in \mathcal{P}(U)} \left[\int_U \int_V e^{\theta r(x,u,v)} \sum_{y \in X} e^{V(\theta,y)} Q(y|x,u,v) \xi(du) \nu^*(x)(dv) \right] \end{cases}$$

- then (μ^*, ν^*) is a pair of saddle point strategies.

References

1. Basu, A., Ghosh, M. K., Zero-sum risk-sensitive stochastic games on a countable state space, *Stochastic Process. Appl.* 124(1), 961-983, 2014.
2. Basu, A., Ghosh, M. K., Nonzero-sum risk-sensitive stochastic games on a countable state space, Submitted.

THANK YOU