

Stochastic games with constraints

新潟大・理学部 田中謙輔 (Kensuke Tanaka)

ハルビン師範大 劉 兆華 (Zhaohua Liu)

制約条件をもつ 2 人零和確率ゲームを次の様に 10 個の要素で与える事にする:

$$(X, Y, A, B, q_1, q_2, r, g^1, g^2, \beta),$$

ただし, X, Y はそれぞれ player 1, player 2 の状態空間, 直積 XY はゲームの状態空間, A, B はそれぞれ player 1, player 2 の行動空間, $q_1(C | x, a)$ は $(x, a) \in XA$ が与えられたときの状態空間 X 上の Borel set C の確率, $q_2(D | y, b)$ は $(y, b) \in YB$ が与えられたときの状態空間 Y 上の Borel set D の確率, $p(CD | (x, a), (y, b)) = q_1(C | x, a)q_2(D | y, b)$ は $((x, a), (y, b)) \in XAYB$ が与えられたときの XY 上の Borel set CD へのゲームの推移確率, $r((x, a), (y, b))$ は実関数: $XAYB \rightarrow \mathbb{R}$ で player 1 の 1-段階の損失関数で $-r$ は player 2 の 1-段階の損失関数, $g^1 = (g_1^1(x, a), g_2^1(x, a), \dots, g_n^1(x, a))$ はベクトル値関数: $XA \rightarrow \mathbb{R}^n$ で player 1 の 1-段階の cost 関数, $g^2 = (g_1^2(y, b), g_2^2(y, b), \dots, g_m^2(y, b))$

はベクトル値関数: $YB \rightarrow R^m$ で player 2 の 1-段階の cost 関数,
 β は割引因子で $0 < \beta < 1$ とする.

ここで, player 1 の strategy $\pi = (\pi_1, \pi_2, \dots, \pi_t, \dots)$ の各要素 π_t は $H_t = H_{t-1}(AX)$ ($H_1 = X$), $t \geq 2$, が与えられたときの A 上の条件付き確率である. 特に, π の全ての要素 π_t が時刻 t に関係しないとき, この strategy は定常と呼ばれる. player 2 の strategy $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_t, \dots)$ の各要素 σ_t は $H_t = H_{t-1}(BY)$ ($H_1 = Y$), $t \geq 2$, が与えられたときの B 上の条件付き確率である. strategy σ に対しても同様に常定性が与えられる.

このとき, player 1 の総期待割引損失は次の様に与えられる:

$$l(\pi, \sigma) = \sum_{t=1}^{\infty} \beta^{t-1} E_{\pi, \sigma} [r((x_t, a_t), (y_t, b_t))].$$

更に, player 1 の総期待割引 cost は次の様に与えられる:

$$G^1(\pi) = \sum_{t=1}^{\infty} \beta^{t-1} E_{\pi} [g^1(x_t, a_t)] \in R^n.$$

亦, player 2 の総期待割引 cost は次の様に与えられる:

$$G^2(\sigma) = \sum_{t=1}^{\infty} \beta^{t-1} E_{\sigma} [g^2(y_t, b_t)] \in R^m.$$

次に, $(u, v) \in R_+^n \times R_+^m$ を用いて, $l(\pi, \sigma)$ を (u, v) の線形関数 $\langle u, G^1(\pi) \rangle + \langle v, G^2(\sigma) \rangle$ で perturb したゲームを考察するために次の様な 2 個の記号を導入することにする:

$$V^+(u, v) = \inf_{\pi} \sup_{\sigma} [I(\pi, \sigma) + \langle u, G^1(\pi) \rangle - \langle v, G^2(\sigma) \rangle],$$

と

$$V^-(u, v) = \sup_{\sigma} \inf_{\pi} [I(\pi, \sigma) + \langle u, G^1(\pi) \rangle - \langle v, G^2(\sigma) \rangle].$$

定理 1 σ_0 が $(\bar{u}, \bar{v}) \in R_+^n \times R_+^m$ の線形関数で perturb した確率ゲームにおける player 2 の max-inf strategy とするとき, 次の事が成立する:

$$-G^2(\sigma_0) \in \partial V^+(\bar{u}, \cdot)(\bar{v})$$

で

$$\inf_{v \in R_+^m} [V^+(\bar{u}, v) + \langle v, G^2(\sigma_0) \rangle] = \inf_{\pi} [I(\pi, \sigma_0) + \langle \bar{u}, G^1(\pi) \rangle].$$

ただし, $\partial V^+(\bar{u}, \cdot)(\bar{v})$ は v での $V^+(\bar{u}, \cdot)$ の subgradient の集合を表す記号で, 全ての $z \in \partial V^+(\bar{u}, \cdot)(\bar{v})$ に対して 不等式

$$V^+(\bar{u}, v) - V^+(\bar{u}, \bar{v}) \geq \langle v - \bar{v}, z \rangle$$

が成立している.

定理 2 π_0 が $(\bar{u}, \bar{v}) \in R_+^n \times R_+^m$ の線形関数で perturb した確率ゲームにおける player 2 の mini-sup strategy とする

とき，次の事が成立する：

$$G^1(\pi_0) \in \partial V^-(\cdot, \bar{v})(\bar{u})$$

で

$$\sup_{u \in R_+^n} [V^-(u, \bar{v}) - \langle u, G^1(\pi_0) \rangle] = \sup_{\sigma} [I(\pi_0, \sigma) - \langle \bar{v}, G^2(\sigma) \rangle]$$

ただし，ここでは全ての $y \in \partial V^-(\cdot, \bar{v})(\bar{u})$ に対して不等式

$$V^-(u, \bar{v}) - V^-(\bar{u}, \bar{v}) \leq \langle u - \bar{u}, y \rangle$$

が成立している。

定理3 X, Y はそれぞれ Polish space の Borel set, A, B はそれぞれ Polish space の compact set, 推移確率 $p = q_1 q_2$ を構成する q_1, q_2 は $(x, y) \in XY$ が与えられたとき，それぞれ A, B 上で連続， r は $XAYB$ 上で連続で有界， g^1 は XA 上で連続で $\|g^1\|$ は有界， g^2 は YB 上で連続で $\|g^2\|$ は有界と仮定すると， $(\bar{u}, \bar{v})$ の線形関数で perturb した確率ゲームに対して player 1 と player 2 の定常な strategy の鞍部点 (π_0, σ_0) とゲームの値：

$$V^+(\bar{u}, \bar{v}) = V^-(\bar{u}, \bar{v}) = V(\bar{u}, \bar{v})$$

が存在する。ただし，

$$V(\bar{u}, \bar{v}) = l(\pi_0, \sigma_0) + \langle \bar{u}, G^1(\pi_0) \rangle - \langle \bar{v}, G^2(\sigma_0) \rangle.$$

定理4 定理3の条件のもとで，次の事が成立する．

$$V(\bar{u}, \bar{v}) + \langle \bar{v}, G^2(\sigma_0) \rangle \leq \sup_{u \in R_+^n} \inf_{\pi} [l(\pi, \sigma_0) + \langle u, G^1(\pi) \rangle]$$

$$= \inf_{G^1(\pi) \leq \theta} l(\pi, \sigma_0)$$

で

$$V(\bar{u}, \bar{v}) - \langle \bar{u}, G^1(\pi_0) \rangle \geq \inf_{v \in R_+^m} \sup_{\sigma} [l(\pi_0, \sigma) - \langle v, G^2(\sigma) \rangle]$$

$$= \sup_{G^2(\sigma) \leq \theta} l(\pi_0, \sigma).$$

定理5 (π_0, σ_0) が制約条件 $G_1(\pi) \leq \theta$, $G^2(\sigma) \leq \theta$ をもつ primal game の鞍部点，即ち

$$\inf_{G^1(\pi) \leq \theta} l(\pi, \sigma_0) = l(\pi_0, \sigma_0) = \sup_{G^2(\sigma) \leq \theta} l(\pi_0, \sigma),$$

とするとき，ある条件のもとで次の式を満たす player 1 の strategy $\bar{\pi}$ と $u^* \in R^n$ が存在する：

$$\begin{aligned} I(\pi_0, \sigma_0) &= \inf_{G^1(\pi) \leq \theta} I(\pi, \sigma_0) \\ &= I(\bar{\pi}, \sigma_0) + \langle u^*, G^1(\bar{\pi}) \rangle. \end{aligned}$$

亦，次の式を満たす player 2 の strategy $\bar{\sigma}$ と $v^* \in R^m$ が存在する：

$$\begin{aligned} I(\pi_0, \sigma_0) &= \sup_{G^2(\sigma) \leq \theta} I(\pi_0, \sigma) \\ &= I(\pi_0, \bar{\sigma}) - \langle v^*, G^2(\bar{\sigma}) \rangle. \end{aligned}$$

更に，ゲームの値について次の関係が成立する：

$$I(\pi_0, \sigma_0) = V^+(u^*, v^*) = V^-(u^*, v^*).$$

参考文献

1. Liu Zhaohua & K.Tanaka, On the closest multistrategy to the shadow minimum of a Markov game, 数理解析研究所講究録, 611, 1987.
2. Liu Zhaohua & K.Tanaka, On an optimal multistrategy and a weak optimal multistrategy of a Markov game, Sci. Rep. Niigata Univ., Vol.23,1987, pp 1-11.
3. K.Tanaka & C.Matsuda, On a continuously discounted vector valued Markov decision process, J.Information & Optimization Science, Vol.11,No.1,1990,pp 33-48.
4. K.Tanaka, On a discounted dynamic programming with the constraints, to appear in JMAA.
5. K.Tanaka & K.Yokoyama, On ε -equilibrium point in a noncooperative n-person game. to appear in JMAA.