2F2-1

私的観測付き繰り返しゲームの三角取引モデルにおける均衡解析

Analysis of Equilibria in Repeated Triangular Trade with Private Monitoring

重冨 孝太 *1 Kota Shigedomi
重中 風奎 *1 Fuuki Shigenaka
岩崎 敦 *2 Atsushi Iwasaki
関口 格 *3 Tadashi Sekiguchi
横尾 真 *1 Makoto Yokoo

*1 九州大学 大学院システム情報科学府
Graduate School of Information Science and Electrical Engineering at Kyushu University

*2 電気通信大学 大学院情報システム学研究科
Graduate School of Information Systems, University of Electro-Communications

*3 京都大学 経済研究所経済戦略研究部門
Strategic Economic Studies Division, Institute of Economic Research, Kyoto University

A repeated game with imperfect private monitoring, where players cannot observe the opponents' actions, is a formal method to analyze long-term relationships. In triangular trade with private monitoring, there exist three players: 0, 1, and 2. Player 0 makes a product and delivers it to player 1. Similarly, player 1 does it for player 2, and player 2 does it for player 0. When player 0 deceives and delivers a low-quality product to player 1, player 1 cannot punish player 0 directly under settings of triangular trade, therefore sustaining cooperation is difficult. We add two new actions and corresponding signals, and identify an equilibrium strategy by which players can sustain cooperation.

1. 序論

無限回繰り返しゲーム (以降繰り返しゲーム) は，プレイヤ間の長期的な協力関係を考察するための枠組みとして知られる．繰り返しゲームにおいては，プレイヤが他のプレイヤの行動を正確に観測できるか否かで問題が二つに大別される．他のプレイヤの行動が完全に観測できる場合は完全観測 (perfect monitoring) と呼ばれ，既に多くの研究成果が得られている．一方，他のプレイヤの行動を，ノイズを含むシグナルを通して不完全にしか観測できない場合は不完全観測 (imperfect monitoring) と呼ばれる．特に，プレイヤが観測するシグナルが私的なもので，そのシグナルを他のプレイヤが観測できない場合は，**不完全私的観測** (imperfect private monitiring) と呼ばれている．不完全私的観測は自然な仮定であり，様々な応用事例が存在する [Mailath 06]．しかし他のプレイヤの選択した行動を完全には知ることができないため，プレイヤ間の協調が困難となる．また，各プレイヤの行動に関しての統計的推測も考慮しなければならず，完全観測と比較して解析することも困難である．

三角取引ではプレイヤ 0，プレイヤ 1，プレイヤ 2 の 3 人のプレイヤが存在し，各期でプレイヤ 0 はプレイヤ 1 に，プレイヤ 1 はプレイヤ 2 に，プレイヤ 2 はプレイヤ 0 に自分の作った製品を提供する．各プレイヤは期間ごとに，製品を作る際に努力するか否かを選択する．努力すると高品質の製品が提供される確率が大きくなり，プレイヤ全員が努力した場合に全員の期待利得の合計を最大化できる．しかし製品を作る際に努力する場合にはコストを要するため，各プレイヤには，努力せずに粗悪品を提供する誘因が生じる．繰り返しゲームの多くのモデルにおいては，裏切りを選択したプレイヤに対して懲罰を与え，裏切ったプレイヤの将来的な利得を減少させる戦略を構築することで，協力関係を維持することが可能である．しかし三角取引においては，プレイヤ 0 がプレイヤ 1 を裏切って粗悪品を提供した場合，プレイヤ 1 はプレイヤ 0 を直接罰することが不可能である．したがって，協力関係を維持することは困難である．

実社会において，2 者が互いに与える影響力が同等でなく，自分を裏切った相手を直接罰することが不可能である関係は数多く存在する．この関係がサイクルを形成している場面も多く，例としてテレビ番組の視聴者，制作会社，スポンサーの 3 者の関係が考えられる．このような場合は全員が高品質のものを循環させることで社会的に望ましい結果を得られる．しかし高品質のものを循環させるにはコストを要するため，個人単位で考えると，裏切る誘因が生じる場合がある．三角取引は，このような状況を簡潔に表現している．

本論文ではプレイヤの戦略を有限オートマトン (FSA: Finite State Automaton) を用いて表現する．プレイヤの行動が努力する (C)，努力しない (D) の 2 通りのみである場合に戦略 FSA が信念不問均衡を構成するか否かを網羅的に判定した．その結果，現実的なパラメータの下で信念不問均衡を構成する状態数 3 以下の戦略 FSA は，全てのピリオドで行動 D を選択する戦略を除き，発見されなかった．そこで本論文では，新たな行動と，各行動に対応するシグナルを追加したモデルを考察する．追加した行動は，相手に罰を与えるための行動と，裏切りが発生したことを知らせ，罰を与えるように依頼するための行動の二つであり，行動 C の自然な拡張により実現可能である．このモデルにおいて協力関係を維持することが可能となる**間接罰則戦略**を提案する．

以下に本論文の構成を記載する．第 2 節では三角取引を私的観測付き繰り返しゲームとして定式化する．第 3 節では戦略と信念不問均衡の定義を述べる．第 4 節では行動を追加したモデルにおける均衡戦略を示し，その理論的分析を行う．第 5 節では本論文の結論を示す．

連絡先: 重冨孝太，九州大学大学院システム情報科学府，819-0395 福岡県福岡市西区元岡 744 番地，(092)802-3576，shigedomi@agent.inf.kyushu-u.ac.jp

2. モデル

本節では本論文で扱うモデルを定める．プレイヤの集合を $N = \{0, 1, 2\}$ で表す．本論文では任意の整数 j に対し，“プレイヤ j” と表記した場合プレイヤ $j \bmod 3$ を表すものとする．無限回繰り返しゲームは，ステージゲームと呼ばれる同一のゲームを無限回繰り返すことを 1 つのゲームとみなしたものである．t 回目のステージゲームをピリオド t $(t = 0, 1, 2, \ldots)$ と呼び，各ピリオドで各プレイヤ $i \in N$ は，プレイヤ $i+1$ に自分が作った製品を提供する．各プレイヤの行動は製品を作る際の努力で表され，行動集合を $A = \{C, D, W, P\}$ で定める．行動 C は製品を作る際に努力すること，行動 D は製品を作る際に努力しないことを意味する．行動 W は製品を作る際に努力しつつ，裏切りの発生を知らせること，行動 P は知らせを受けて相手に罰を与えることを意味する．行動 W, P はともに相手に与える便益や情報に関する変化を C に対して加えたものであり，行動 C の自然な拡張により実現可能である．プレイヤ $i-1$ が各行動 C, D, W, P を選択した場合にプレイヤ i が受け取る期待便益をそれぞれ $1, 0, 1-y, 1-z$ と定める．すなわち，プレイヤ $i-1$ が行動 W, P を選択した場合にプレイヤ i が受け取る期待便益は，行動 C による期待便益と比較してそれぞれ y, z の損失を受ける．また，各行動を選択した場合にプレイヤが支払うコストを，行動 D は 0，D 以外の行動は c と定める．

各プレイヤが受け取るシグナルは受け取った製品の質で表され，シグナル集合を $\Omega = \{g, b, w, p\}$ で定める．シグナル g は受け取った製品が高品質であったこと，シグナル b は受け取った製品が低品質であったことを意味する．シグナル w は高品質の製品を受け取りつつ裏切りの発生を知ったこと，シグナル p は罰を与えられたことを意味する．シグナル g, b, w, p はそれぞれ行動 C, D, W, P に対応する正しい (エラーのない) シグナルである．ここで，各プレイヤの行動と受け取ったシグナルは私的である，すなわち他のプレイヤの行動と受け取ったシグナルを直接観測することができないという仮定を置く．プレイヤ i の行動を a_i で表し [*1]，プレイヤ i がプレイヤ $i-1$ から受け取るシグナルを ω_i で表す．ここで，各プレイヤの受け取るシグナルは独立であり，プレイヤ i が受け取るシグナルの確率分布は，プレイヤ $i-1$ の行動のみに依存すると仮定する．また任意の a_{i-1} に対し，いずれのシグナル ω_i も正の確率で発生することを仮定する (フルサポートの仮定)．プレイヤ $i-1$ が行動 a_{i-1} を選択した場合にプレイヤ i がシグナル ω_i を受け取る確率 $o(\omega_i \mid a_{i-1})$ を以下で表す．

$$o(\omega_i \mid a_{i-1}) = \begin{cases} q & (\text{正しいシグナル}) \\ e = (1-q)/3 & (\text{誤ったシグナル}) \end{cases}$$

ただし，$1/4 < q < 1$，すなわち正しいシグナルが伝わる確率が最も高いと仮定する．

各プレイヤ i の実現利得は，自身の行動と受け取ったシグナルによって決定する利得であり，$\pi_i(a_i, \omega_i)$ と表される．一方，プレイヤ i の期待利得は $\sum_{\omega_i \in \Omega} \pi_i(a_i, \omega_i) \cdot o(\omega_i \mid a_{i-1})$ で表され，この値をステージゲーム利得 $u_i(\boldsymbol{a})$ として扱う．ただし $\boldsymbol{a} = (a_0, a_1, a_2)$ である．これらの設定は，各ピリオドで獲得した利得から相手の行動を観測することが不可能であることを意味する．各プレイヤの利得は受け取った便益と支払ったコストの差で表されるため，プレイヤ i のステージゲーム利得

*1 任意の整数 j に対し a_j でプレイヤ $j \bmod 3$ の行動を表す．

表 1: プレイヤ i のステージゲーム利得

$a_i \backslash a_{i-1}$	C	D	W	P
C	$1-c$	$-c$	$1-y-c$	$1-z-c$
D	1	0	$1-y$	$1-z$
W	$1-c$	$-c$	$1-y-c$	$1-z-c$
P	$1-c$	$-c$	$1-y-c$	$1-z-c$

は表 1 の通りである．また，以上の理由から，プレイヤ $i+1$ の行動はプレイヤ i のステージゲーム利得に影響を与えない．

ここで，パラメータに制約を与える．まず $y, z \geq 0$，$0 < c < 1$ とする．この制約は，全員が行動 C を選択している状態が効率的，すなわちプレイヤ全員の利得の和を最大化することを表す．また，$z \leq 1$ とする．この制約は，行動 P によって相手に与える便益が行動 D による便益より小さくなるような強い罰を考えないことを表す．以上の制約は，本論文で扱うモデルが $A = \{C, D\}$，$\Omega = \{g, b\}$ であるモデルの自然な拡張となるようにするためにも必要である．ステージゲームを 1 回のみ行うゲームにおいて全員が行動 D を選択することが唯一のナッシュ均衡であるという点は，$A = \{C, D\}$ である三角取引のモデルと本論文で扱うモデルで同一である．

プレイヤ i の割引平均利得を $(1-\delta)\sum_{t=0}^{\infty} \delta^t u_i(\boldsymbol{a}^t)$ として定義する．これは各ピリオドで与えられる利得の平均値を表す [*2]．ただし，$\boldsymbol{a}^t = (a_0^t, a_1^t, a_2^t)$ はピリオド t における各プレイヤの行動を，$\delta \in (0, 1)$ は割引因子 (将来の利得を重視する程度) をそれぞれ表す．各プレイヤは各ピリオドにおいて，自身の割引平均利得を最大化するように行動を選択すると仮定する．

3. 戦略の表現と均衡概念

本節では，繰り返しゲームにおける戦略を定義し，FSA を用いて戦略を表現する方法を述べる．また，均衡概念の一つである信念不問均衡に関しても述べる．

プレイヤ $i \in N$ のピリオド t における私的ヒストリの集合を，$H_i^t := (A \times \Omega)^t$ とする．$h_i^t \in H_i^t$ を，ピリオド $t-1$ までの各ピリオドにおけるプレイヤ i の選択した行動と観測したシグナルを記述したものとする．また，h_i^0 を，ピリオド $t = 0$ における行動を決定するためのダミー変数とし，$H_i^0 := \{h_i^0\}$ と定める．プレイヤ i の (純粋) 戦略は，関数 $s_i : H_i \to A$ によって表される．ただし $H_i := \bigcup_{t \geq 0} H_i^t$ とする．戦略プロファイルを $\boldsymbol{s} = (s_i, \boldsymbol{s}_{-i})$ と表す．ただし $\boldsymbol{s}_{-i}$ はプレイヤ i 以外の戦略の組を表す．$E_i(\boldsymbol{s})$ を，戦略プロファイル $\boldsymbol{s}$ におけるプレイヤ i の割引平均利得とする．プレイヤ i の戦略 s_i が $\boldsymbol{s}_{-i}$ に対する**最適反応**であるとは，プレイヤ i の任意の戦略 s_i' に対して $E_i((s_i, \boldsymbol{s}_{-i})) \geq E_i((s_i', \boldsymbol{s}_{-i}))$ であることを表す．

プレイヤ $i \in N$ の戦略 FSA を $M_i = \langle \Theta_i, \hat{\theta}_i, f_i, T_i \rangle$ により定義する．Θ_i は状態の集合を表す．$\hat{\theta}_i \in \Theta_i$ は初期状態を表す．$f_i : \Theta_i \to A$ は 各状態においてプレイヤが選択する行動を表す．$T_i : \Theta_i \times \Omega \to \Theta_i$ は各状態におけるシグナルによる決定的遷移を表す．プレイヤ i の戦略プレオートマトン (pre-FSA: Finite State preAutomaton) m_i を $m_i = \langle \Theta_i, f_i, T_i \rangle$ で定義すると，$M_i = (m_i, \hat{\theta}_i)$ で表される．また，戦略 FSA プロファイルを $\boldsymbol{M} = (M_0, M_1, M_2)$ と定める．

*2 プレイヤ i が各ピリオドで与えられる利得の平均値を x_i とすると，割引利得の合計は $x_i + \delta x_i + \delta^2 x_i + \cdots = \frac{x_i}{1-\delta} = \sum_{t=0}^{\infty} \delta^t u_i(\boldsymbol{a}^t)$ であることから言える．

表 2: 直前に受け取ったシグナルと行動の対応

シグナル	行動
g	C
b	W
w	P
p	C

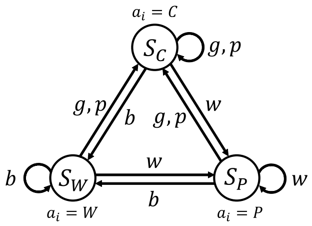

図 1: 間接罰則戦略の戦略 pre-FSA

また，本論文では**信念不問均衡** [Piccione 02, Ely 02] と呼ばれる均衡概念を考える．信念不問均衡は以下で定義される．

定義 1 (信念不問均衡) $\boldsymbol{M} = ((m_0, \hat{\theta_0}), (m_1, \hat{\theta_1}), (m_2, \hat{\theta_2}))$ が信念不問均衡であるとは，すべての t, $(\theta_0, \theta_1, \theta_2) \in \Theta_0^t \times \Theta_1^t \times \Theta_2^t$, $i \in N$ に対して，すべてのプレイヤ $j \in N \backslash \{i\}$ が (m_j, θ_j) に従って行動を選択する場合に (m_i, θ_i) が最適反応となることを意味する．ただし Θ_i^t は，M_i に対しピリオド t で到達可能な状態の集合を表す．

プレイヤ i の信念は，i 以外のプレイヤの私的ヒストリに関する確率的推定を指す．信念不問均衡である戦略は，信念を用いずにプレイヤの最適反応戦略を決定するため，戦略の実行に信念を計算する必要がない．したがって解析を行うことが簡単になる．ここで注意すべき点は，信念不問均衡の定義において，プレイヤの取りうる戦略に制限を加えていないことである．すなわち，FSA で表せない複雑な戦略に逸脱しても，自身の利得を増加させることは不可能であることを意味する．

4. 間接罰則戦略

本節では，与えたモデルにおいて協力関係を維持する均衡を構成する**間接罰則戦略**に関して述べる．間接罰則戦略に従っているプレイヤは最初のピリオドでは行動 C を選択し，以降のピリオドでは，直前のピリオドで受け取ったシグナルのみによって自分の行動を決定する．直前に受け取ったシグナルと自分の行動の対応は表 2 の通りである．戦略 pre-FSA は図 1 で表され，初期状態は S_C である．

間接罰則戦略において裏切ったプレイヤが罰せられる流れを述べる．まず全員が状態 S_C から開始し，全員が行動 C を選択する．ここでプレイヤ 0 がピリオド t_0 で裏切って行動 D を選択したとする．この場合，ピリオド t_0 でプレイヤ 1 は高い確率でシグナル b を受け取る．シグナル b を受け取ったプレイヤ 1 はピリオド $t_0 + 1$ で行動 W を選択し，プレイヤ 2 に裏切りの発生を知らせる．プレイヤ 2 は高い確率でシグナル w を受け取り，ピリオド $t_0 + 2$ で行動 P を選択してプレイヤ 0 を罰する．したがってプレイヤ 0 は行動 D を選択したことで，2 ピリオド後に高い確率でプレイヤ 2 に罰せられる．

裏切ったプレイヤは以上の方法で罰せられ，処罰の流れが終わるとプレイヤは協力 (状態 S_C) に戻る．ゆえに間接罰則戦略によってプレイヤの協力関係が維持可能である．

4.1 均衡構成の条件

本節では，プレイヤ全員が間接罰則戦略を取っている状態が均衡を構成するための条件を示し，その証明を行う．次の定理は，間接罰則戦略の組が信念不問均衡を構成するための条件を示す．証明は節の最後に述べる．

定理 1 (間接罰則戦略の均衡構成条件) 間接罰則戦略の組が信念不問均衡を構成するための必要十分条件は以下の式で表される．

$$\delta^2(1-4e)^2 z \geq c \tag{1}$$

パラメータと協力関係の維持しやすさの関係性に関して直感的に以下のことが言える．まず，努力するために必要なコスト c が小さく割引因子 δ が大きい，すなわちプレイヤが将来の利得を重視するほど協力関係が維持しやすい．また，誤ったシグナルが伝わる確率 e が小さく，相手に与える罰の強さを表す z の値が大きいほど，協力関係が維持しやすい．これらは定理 1 の内容と一致している．

以下に定理 1 の証明の方針を示す．本論文では全員が同じ戦略に従う場合のみを考えるため，一般性を失わずにプレイヤ 0 の最適性のみを考えることができる．プレイヤ 0 が間接罰則戦略から逸脱して行動を選択する誘因が生じないことを示す．戦略からの逸脱としては多くの戦略が考えられるが，文献 [Mailath 06] によると，1 ピリオドのみ戦略から逸脱した行動を選択し，その後は元の戦略に戻って行動を選択する (**一回逸脱**) 場合のみを考えればよい．このことは "**一回逸脱の原則**" と呼ばれる．

プレイヤ $i \in N$ の戦略 s_i に対する一回逸脱戦略 s_i^{dev} を，プレイヤ i が特定の私的ヒストリ $\tilde{h}_i^t$ においてのみ s_i と異なる行動を選択し，その他のピリオドでは s_i に従う戦略とする．プレイヤ i の戦略 s_i に対する一回逸脱戦略の集合を $Q_i(s_i)$ とする．一回逸脱の原則は以下の命題により表される．

命題 1 (一回逸脱の原則) プレイヤ i の戦略 s_i に対するすべての一回逸脱戦略 $s_i^{\text{dev}} \in Q_i(s_i)$ について $E_i((s_i, \boldsymbol{s}_{-i})) \geq E_i((s_i^{\text{dev}}, \boldsymbol{s}_{-i}))$ が成り立つ場合，かつその場合に限り，プレイヤ i の戦略 s_i は $\boldsymbol{s}_{-i}$ に対する最適反応である．

定理 1 の証明には，以下の三つの補題を用いる．証明は紙幅の都合上省略する．

補題 1 (現状態による状態遷移確率の不変性) ピリオド t におけるプレイヤ i の状態に関する確率分布を $\boldsymbol{\gamma}_{i,t} = (\gamma_{S_C} \quad \gamma_{S_P} \quad \gamma_{S_W})$ で表す．ただし γ_{θ_i} は，プレイヤ i の状態が θ_i である確率を表す．また，$\boldsymbol{\gamma}_{i,t}$ と $k \in \mathbb{N}$ に対し，ピリオド $t + k$ におけるプレイヤ $i + k$ の状態に関する確率分布を $\boldsymbol{\gamma}_{i+k,t+k}$ と定める．ただし，プレイヤは全員間接罰則戦略に従うものとする．このとき，すべての $i \in N$, $k \geq 2$, $\boldsymbol{\gamma}_{i,t}, \boldsymbol{\gamma}'_{i,t}$ に対し，以下の式が成り立つ．

$$\boldsymbol{\gamma}_{i+k,t+k} = \boldsymbol{\gamma}'_{i+k,t+k}$$

補題 2 (D 以外の行動に関する信念不問性) $V^{(\theta_0, \theta_1, \theta_2)}$ を，各プレイヤの状態が $(\theta_0, \theta_1, \theta_2)$ であるときのプレイヤ 0 の割引平均利得とする．このとき，すべての $\theta_1, \theta_2 \in \{S_C, S_W, S_P\}$ に対して以下の式が成り立つ．

$$V^{(S_C, \theta_1, \theta_2)} = V^{(S_W, \theta_1, \theta_2)} = V^{(S_P, \theta_1, \theta_2)}$$

補題 3 (現行動による状態遷移確率の不変性) ピリオド t で，プレイヤ i が行動 $a_i \in A$ を選択したと仮定する．ただしプ

レイヤ i はピリオド t で間接罰則戦略に従わずに行動を選択してもよく，その後は間接罰則戦略に従うものとする．$k \geq 3$ であるとき，ピリオド $t+k$ におけるプレイヤ $i+k$ の状態に関する確率分布はすべての a_i で同一である．

以上の補題をもとに，定理 1 の証明を行う．

証明 (定理 1) 命題 1 より，各プレイヤに自身の利得を増加させる一回逸脱戦略が存在しないことが，戦略の組が信念不問均衡を構成するための必要十分条件となる．一回逸脱によってプレイヤ 0 の利得が増加しないための条件の導出を行う．

プレイヤ 0 の一回逸脱を考える際には，逸脱したピリオドでプレイヤ 0 が支払うコストと，逸脱した 2 ピリオド後にプレイヤ 2 から受ける期待便益のみを考えればよい．最初に，$(\theta_0, \theta_1, \theta_2) = (S_C, S_C, S_C)$ の場合に行動を C から D へ変える一回逸脱を考える．逸脱したピリオドでは，D を選択した場合はコストを支払わないため利得は c 増加する．また，行列 X を以下で定義する．

$$X = \begin{pmatrix} 1-2e & e & e \\ 1-2e & e & e \\ 2e & 1-3e & e \end{pmatrix}$$

逸脱した 2 ピリオド後におけるプレイヤ 2 の状態に関する確率分布は，プレイヤ 0 が C を選択した場合は $(1-2e \quad e \quad e)\,X$，プレイヤ 0 が D を選択した場合は $(2e \quad e \quad 1-3e)\,X$ で表される．したがってピリオド t_0+2 にプレイヤ 0 が得る割引期待利得の差は以下で表される．

$$\begin{aligned}
&\delta^2\,(2e \quad e \quad 1-3e)\,X\,(1 \quad 1-z \quad 1-y)^{\mathrm{T}} \\
&\quad -\delta^2\,(1-2e \quad e \quad e)\,X\,(1 \quad 1-z \quad 1-y)^{\mathrm{T}} \\
&= \delta^2\,(-(1-4e) \quad 0 \quad 1-4e)\,X\,(1 \quad 1-z \quad 1-y)^{\mathrm{T}} \\
&= -\delta^2(1-4e)^2 z
\end{aligned}$$

ゆえに一回逸脱によって利得が増加しないための条件は式 (1) により表される．

$\theta_0 = S_W, S_P$ の場合に行動を D に変える場合は，$\theta_0 = S_C$ と同様の結果が導かれる．また，C, W, P 間での逸脱に関しては，補題 2 より逸脱した場合でも利得が増加しない．以上の議論は他プレイヤの状態に依存しないため，$(\theta_1, \theta_2) = (S_C, S_C)$ 以外の場合も同様の結果が導かれる．ゆえに式 (1) は，間接罰則戦略の組が信念不問均衡を構成するための必要十分条件である． □

4.2 利得

本節では，間接罰則戦略においてプレイヤに与えられる利得の分析を行う．以下の定理は，プレイヤ全員が間接罰則戦略に従っている場合の各プレイヤの割引平均利得を表す．

定理 2 (間接罰則戦略によって与えられる利得) プレイヤ全員が間接罰則戦略に従っている場合，各プレイヤの割引平均利得は以下の式で表される．

$$(1-c) - \delta^2 e(1-4e)z - \delta e y - \delta e z \tag{2}$$

証明 プレイヤ全員が間接罰則戦略に従っている限りプレイヤは行動 D を選択しないため，支払うコストは全ピリオドで c である．一方，プレイヤ全員が間接罰則戦略に従い，かつ，シグナルにエラーが生じない場合に各プレイヤが受け取る便益は 1 である．シグナルにエラーが生じ，他プレイヤが行動 W, P を選択することによる，割引平均利得の減少分は以下で表される．ただし 2 行目の等号は，すべての $k \geq 2$ に対し $X^k = X^2$ であることより成立する．

$$\begin{aligned}
&(1-\delta)\sum_{t=0}^{\infty}\delta^t\,(1 \quad 0 \quad 0)\,X^t\,(0 \quad z \quad y)^{\mathrm{T}} \\
&= (1-\delta)\delta\,(1-2e \quad e \quad e)\,(0 \quad z \quad y)^{\mathrm{T}} \\
&\quad + \delta^2\left(1-3e+4e^2 \quad 2e-4e^2 \quad e\right)(0 \quad z \quad y)^{\mathrm{T}} \\
&= \delta e y + \delta e z + \delta^2 e(1-4e)z
\end{aligned}$$

ゆえに各プレイヤの割引平均利得は式 (2) により表される． □

定理 1 と 定理 2 より，y の値は利得には関与しているが，均衡構成条件には関与していない．この直感的な理由を以下に述べる．間接罰則戦略に従う限り行動 D は選択されないが，シグナルのエラーにより，プレイヤがシグナル b を受け取る可能性はある．シグナル b を受け取ったプレイヤは次ピリオドで行動 W を選択するため，行動 W による便益に関わる値 y は利得に関与する．一方，プレイヤ全員が間接罰則戦略に従っている場合にプレイヤ $i-1$ が行動 W を選択するのは，直前のピリオドでプレイヤ $i+1$ からシグナル b が伝わった場合のみである．プレイヤ全員が間接罰則戦略に従う限り，プレイヤ $i+1$ からプレイヤ $i-1$ にシグナル b が伝わる確率は変化しない．したがって y の値が変化しても均衡戦略からの逸脱による利益は変化しないため，y の値は均衡構成条件には関与しない．

5. 結論

本論文では，私的観測付き繰り返しゲームの三角取引モデルにおいて，行動と対応するシグナルを追加したモデルの分析を行った．このモデルで協力関係を維持できる均衡戦略である間接罰則戦略を提示し，間接罰則戦略の組が信念不問均衡を構成する条件と，各プレイヤに与えられる利得を導出した．間接罰則戦略は，各ピリオドで選択する行動が直前のピリオドで受け取ったシグナルのみに依存するため，行動の選択における計算量の観点から簡単な戦略であると言える．今後の課題としては，間接罰則戦略によって与えられる利得の最適性の検証が挙げられる．

謝辞

本研究は JSPS 基盤研究 (S) (24220003) の助成を受けました．深く感謝致します．

参考文献

[Ely 02] Ely, J. C. and Välimäki, J.: A Robust Folk Theorem for the Prisoner's Dilemma, *Journal of Economic Theory*, Vol. 102, No. 1, pp. 84–105 (2002)

[Mailath 06] Mailath, G. J. and Samuelson, L.: *Repeated Games and Reputations*, Oxford University Press (2006)

[Piccione 02] Piccione, M.: The Repeated Prisoner's Dilemma with Imperfect Private Monitoring, *Journal of Economic Theory*, Vol. 102, No. 1, pp. 70–83 (2002)