Comparison of information in zero-sum stochastic games

Marcin Pęski, Juuso Toikka

University of Toronto, MIT

05/16/14
Introduction
Zero-sum stochastic

- $g(a, b, s)$ - maximizer’s payoffs,
  - $a \in A$ maximizer’s action
  - $b \in B$ minimizer’s action
  - $s$ - state, $s_{t+1} \sim P(s_t) \in \Delta S$, where $P$ is Markov operator, exogenous,
    - ergodic distribution $\pi_p = P\pi_p \in \Delta S$,
  - maximizer (but not minimizer) observes the state of the world in each period,
- Value $v^\delta(p; g, P)$ for discounting with $\delta < 1$,
  - the limit value (it exists and it does not depend on $p$!):
  $$v(g, P) = \lim_{\delta \to 1} v^\delta(p; g, P)$$
Definition

Operator $Q$ is better for maximizer than $P$ (i.e., $P \preceq Q$) if for each game $g$,

$$v(g, P) \leq v(g, Q).$$

Problem

Characterize relation $P \preceq Q$. 
Introduction

Motivation

- Stochastic games vs. repeated games with incomplete information (i.e., Aumann-Maschler)
  - trade-off between short term gains from using private information and long-term costs of revealing it,
  - there are no short term gains when states are permanent.
- Comparison of information literature: (Blackwell 1953, Mertens-Gossner 01, Peski 08).
  - intuition: more information (in the Blackwell sense) is better for the minimizer,
  - here: more information means that $P$ is more persistent.
  - however, it is difficult to separate the information and the payoff effects of transitions.
- Applications:
  - zero-sum stochastic games (where value is notoriously difficult to compute (Hörner at al, 2010),
  - individual rationality constraint in repeated games,
  - one long-run vs. many short run players.
Two states $s_1, s_2$

- $s_t = s_{t+1}$ with prob. $\rho \in [0, 1]$,
- the larger $\rho$, the more persistent is the state.

Maximizer chooses $U$ or $D$ and the payoffs are described above.

It is notoriously difficult to compute value (Hörner at al, 2010,
Introduction

Plan

1. Introduction
2. Notations and definitions
3. Static formula for the value
4. Characterization (necessary and sufficient) of the order $\preceq$
5. Corollaries
Notations and Definitions

Beliefs

- $p, q \in \Delta S$ - space of (minimizer’s) beliefs,
  - prior beliefs in period $t$: beliefs before the actions are chosen (and information revealed),
  - posterior beliefs in period $t$: beliefs after the actions are chosen,
- If $p$ are posterior beliefs today, then $Pp$ are prior beliefs tomorrow,

\[(Pp)(s) = \sum_{s'} p(s) P(s|s').\]

- $\mu, \psi \in \Delta^2 S = \Delta(\Delta S)$ - distributions over beliefs,
  - for each measurable $A \subseteq \Delta S$
    \[(P\mu)(A) = \mu \{ q : Pq \in A \} = \mu (P^{-1}A).\]
Maximizer strategy \( \alpha : S \to \Delta A \) induces

- posterior beliefs
  \[ p^a(s) = \frac{p(s)\alpha(a|s)}{\sum_{s'} p(s')\alpha(a|s')} \]
  and

- distribution \( m^\alpha \in \Delta^2 S \) over posterior beliefs
  \[ m^\alpha(p^a) = \sum_{s'} p(s') \alpha(a|s') \] for each \( a \).
Notations and Definitions

Distributions over beliefs I: revelation of information

Revelation strategy

Induced posteriors
Zero-sum games

- continuation value depends only on beliefs,
- it is concave in beliefs,

Given $\beta \in \Delta B$, strategy $\alpha : S \rightarrow \Delta A$ is max. best response only if each $a$ played with positive probability:

$$g(a, \beta, p^a) = \max_{a' \in A} g(a', \beta, p^a) =: \hat{g}(\beta, p).$$

If $m \in \Delta^2 S$ is the eq. revelation policy, then the expected payoff is

$$\min_{\beta \in \Delta B} \int \hat{g}(\beta, q) \, dm(q).$$

W.l.o.g. strategies are Markov in priors.
**Definition**

\( \psi \) is a *mean preserving spread* of \( \mu \) if there exists a measurable \( m : \Delta S \to \Delta^2 S \) such that

\[
(i) \quad E m (.|q) = q \text{ for each } q, \text{ and} \\
(ii) \quad \psi(dp) = \int m(dp|q) d\mu(q).
\]

We write \( \psi = \mu \ast m \).

We say that \( \mu \) is a *Blackwell garbling* of \( \psi \), or \( \mu \leq^B \psi \).

- Note that if \( \mu \leq^B \psi \), then \( P\mu \leq^B P\psi \).
- Blackwell: If \( \mu \leq^B \psi \), then there exists concave \( f \) such that \( \mu[f] > \psi[f] \).
ψ is a mean preserving spread of \( \mu \).
Characterization of value
Distributions over beliefs $\mu$: Operator $P$

- Zero-sum stochastic game:
  - w.l.o.g. max. strategy is stationary $\sigma : \Delta S \times S \to \Delta A$,
  - induces m.p.s. $m^\sigma(\cdot)$, and
  - stationary distribution over prior beliefs $\mu$,

- if $\mu$ is a distribution over priors, then
  - $\mu \ast m^\sigma(\cdot)$ is a distribution over posteriors, and
  - $P\left(\mu \ast m^\sigma(\cdot)\right)$ is a distribution over prior beliefs in the next period.

- Because $\mu$ is stationary:
  $$P\left(\mu \ast m^\sigma(\cdot)\right) = \mu.$$
Main Result

**Theorem**

*(Value of the stochastic zero-sum game)* For each $g$,

$$v(g, P) = \max_{\mu, m \text{ st. } P(\mu \ast m) \leq B \mu} \int \left( \min_{\beta \in \Delta B} \int (\hat{g}(\beta, q)) \, dm(q|p) \right) \, d\mu(p)$$

- when $\delta \to 1$, the value converges to the average revelation payoff over the stationary distribution.
Main Result

**Theorem**

(a) $P \preceq Q$ iff for each $\mu$, $m$ st. $P(\mu * m) \leq^K \mu$, we have $Q(\mu * m) \leq^K \mu$,

(b) $P \preceq Q$ iff for each $\mu$ such that $P\mu \leq^K \mu$, we have $Q\mu \leq^K P\mu$.

**Proof:**

- equivalence: take $\mu^{(a)} = P\mu^{(b)}$,
- one direction follows from the characterization of the value.

**Fixed point-ish flavor:** not very easy to use in applications, but easy to use in the proofs.
Main Result: Proof

- Suppose that $P(\mu_0 \ast m_0) \leq^B \mu_0$ and $Q(\mu_0 \ast m_0) \not\leq^B \mu_0$.
- Blackwell: there exists a concave function $f : \Delta S \to R$ st.

\[ \forall_{(\mu, m) \text{ st.}} Q(\mu \ast m) \leq^B \mu \mu[f] - Q(\mu \ast m)[f] < 0 \text{ and } , \]
\[ \mu_0[f] - Q(\mu_0 \ast m_0)[f] > 0. \]

- W.l.o.g. there is a finite set $L$ of functions $l : S \to R$.

\[ f(p) = \min_{l \in L} \sum p(s) l(s), \]
Main Result: Proof

Let $A = B = L$, and for each $a, b \in L$,

$$g(a, b, s) = b(s) - \sum_{s'} Q(s'|s) a(s').$$

We show that

$$\int \left( \min_{\beta \in \Delta B} \int (g^*(\beta, q)) \, dm(q|p) \right) \, d\mu(p) = \mu[f] - Q(\mu * m)[f].$$
Main Result: Proof

We have

\[ g^* (\beta, q) = \max_\alpha \sum_s q(s) g(a, \beta, s) \]

\[ = \sum_s \beta(s) q(s) - \min_\alpha \sum_s q(s) \sum_{s'} Q(s'|s) a(s) \]

\[ = \sum_s \beta(s) q(s) - f(Qq), \]

and

\[ \min_{\beta \in \Delta B} \left( \int g^* (\beta, q) \, dm(q|p) \right) \]

\[ = \left( \min_{\beta \in \Delta B} \sum_s \beta(s) p(s) \right) - \left( \int f(Qq) \, dm(q|p) \right) \]

\[ = f(p) - \left( \int f(Qq) \, dm(q|p) \right). \]
If $P \preceq Q$ or $Q \preceq P$, then $\pi_P = \pi_Q$. 

Corollary
Other observations
Simple (but not complete) characterization

- \( A = \{ (\alpha_1, \alpha_2, ..., \alpha_\infty) : \alpha_i \geq 0, \sum \alpha_i = 1 \} \),
- for each \( \alpha \in A \), let \( P^\alpha = \alpha_1 P + \alpha_2 P^2 + ... + \alpha_\infty P^\infty \).

**Theorem**

For each ergodic \( P, Q \):

1. **Sufficient condition:** If \( Q = P^\alpha \) for some \( \alpha \in A \), then \( P \preceq Q \).
2. **Necessary condition:** If \( P \preceq Q \), then, for each \( p \), there exists \( \alpha_p \in A \) such that \( Qp = P^{\alpha_p} p \).
3. If \( P \) has purely real eigenvalues, then the necessary and the sufficient conditions are equivalent.
Other observations

Operator $P : \Delta S \rightarrow \Delta S$

Real eigenvalues

Complex eigenvalues

Action of operator $P$
Other observations
Simple (but not complete) characterization

- Complete characterization within subspace of operators with real eigenvalues,
- For general operators, we know the sufficient is not necessary.
  - we do NOT know whether the necessary condition is sufficient
  - but the necessary condition is not really easier than our full characterization).
- Proof shows that $P^n \preceq P^{n+1}$.
  - transitions $P$ is more persistent than transitions $P^n$, which are more persistent than $P^{n+1}$,
  - more persistence is good for the minimizer,
  - but this is a very special sense.
Suppose that $|S| = 2$. Then, each $P$ has unique (and real) eigenvalue $\lambda_P \in (-1, 1)$. Moreover, $P \preceq Q$ if and only if $\pi_P = \pi_Q$, and

- if $\lambda_P \geq 0$, then $\lambda_Q \in [0, \lambda_P]$,
- if $\lambda_P \leq 0$, then $\lambda_Q \in [\lambda_P, \lambda_P^2]$.

Apply the necessary and sufficient condition.

Application: monotonicity of value in (Hörner at al, 2010) (for all payoffs)
Other observations

Corollaries

Corollary

If $P \preceq Q$ and $Q \preceq P$, then $P = Q$. So,

Proof.

For each $p \in \Delta S$, let $A^P(p) = \text{con}\{Pp, P^2p, \ldots\}$.

- If $P \preceq Q$, then necessary condition implies $Qp \in A^P(p)$.
- It follows that $A^Q(p) \subseteq A^P(p)$ with strict inclusion if $Qp \neq Pp$.
- Similarly, if $Q \preceq P$ and $Qp \neq Pp$ then $A^P(p) \subsetneq A^Q(p)$. Contradiction.
Other observations
Corollaries

- Let $D_\pi \in \mathcal{P}$ be i.i.d. draw from distribution $\pi \in \Delta S$.

**Corollary**

$P \preceq D_{\pi_P}$.

- It follows from the sufficient condition.
- The opposite result does not hold.
Other observations

Corollaries

Corollary

The set \( \{ Q : P \preceq Q \} \) is convex (in the operator sense).

Proof.

It follows almost immediately from the general characterization.

- For purely real eigenvalues, set \( \{ Q : P \preceq Q \} \) is spanned by \( P^n s \).
- In general case, we don’t know the extreme points of \( \{ Q : P \preceq Q \} \).
Corollary

For each $\alpha \in [0, 1]$, $P \preceq \alpha P + (1 - \alpha) D_{\pi_P}$.

- More persistent information is good for minimizer.
Other observations

Corollaries

- The value of a game with permanent incomplete information

\[ v^\delta (\pi_P, g) := v^\delta (\pi_P; g, l). \]

Corollary

If \( P \neq \alpha I + (1 - \alpha) D_\pi \) for some \( \alpha \in (0, 1) \) and \( \pi \in \Delta S \), then, there exists game \( g \) such that

\[ \liminf_{\delta \to 1} v^\delta (\pi_P, g) > v (g, P). \]
Conclusions

- We analyze stochastic games with incomplete information.
  - formula for the value,
  - various conditions for the comparison of the value with respect to the stochastic process
  - more persistence (in very special sense) is good for the minimizer,