Mean-field games and stochastic games

Josu Doncel Nicolas Gast Bruno Gaujal

COSMOS, July 6, 2016

General Concepts

Mean field games

Controlled dynamics of a large number of particles/players in mean field interaction.

- Interacting particles/players
  Players interact and can take actions to optimize their cost function.
- Notions of solutions
  All players cannot minimize their costs (0-sum system)

  Equilibrium: set of strategies for all players where no player has incentive to change its action.

  Optimal centralized control: minimize the cost of the community (e.g. sum of the costs).

Discrete state space

Mean field games (Lions and Lasry, 2007 and Caines, 2007) capture the dynamic evolution of a large population of strategic players.

Discrete state space

Mean field games (Lions and Lasry, 2007 and Caines, 2007) capture the dynamic evolution of a large population of strategic players.

Here, focus on discrete state space.

Fokker Planck Equation $\Longrightarrow$ Kolmogorov Equation.
Hamilton Jabobi Bellman Eq. $\Longrightarrow$ Bellman optimality Eq.

Stochastic Games Shapley, 1953

Generalization of Markov Decision Processes with multiple synchronous controllers (players), each with her own cost function.

State (configuration) at t:

$\mathbf{X}(t) = (X_1(t), \ldots, X_N(t))$,
with $X_n(t) \in \mathcal{S}$ (finite set).

Cost	0	0	0	0	0	0	0	0	0	0
State	0	1	1	1	0	1	0	0	1	0

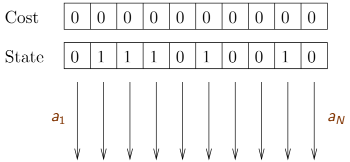
Stochastic Games Shapley, 1953

Generalization of Markov Decision Processes with multiple synchronous controllers (players), each with her own cost function.

State (configuration) at t:

$\mathbf{X}(t) = (X_1(t), \ldots, X_N(t))$,
with $X_n(t) \in \mathcal{S}$ (finite set).

Player n chooses action a_n.
State change follows $P_{ij}(a_n, \mathbf{X})$



Stochastic Games Shapley, 1953

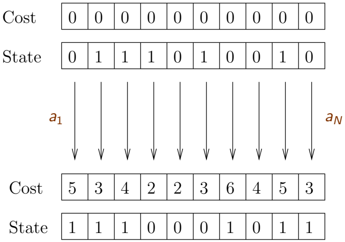
Generalization of Markov Decision Processes with multiple synchronous controllers (players), each with her own cost function.

State (configuration) at t:

$\mathbf{X}(t) = (X_1(t), \ldots, X_N(t))$,
with $X_n(t) \in \mathcal{S}$ (finite set).

Player n chooses action a_n.
State change follows $P_{ij}(a_n, \mathbf{X})$

Her cost increases by
$\beta^t C_n(X_n, a_n, \mathbf{X})$.



Stochastic Games Shapley, 1953

Generalization of Markov Decision Processes with multiple synchronous controllers (players), each with her own cost function.

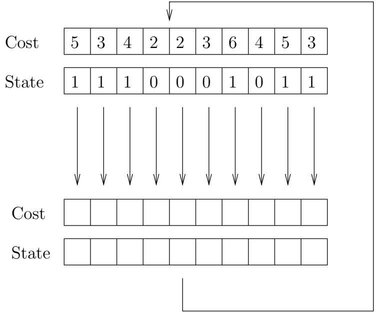
State (configuration) at t:

$\mathbf{X}(t) = (X_1(t), \ldots, X_N(t))$,
with $X_n(t) \in \mathcal{S}$ (finite set).

Player n chooses action a_n.
State change follows $P_{ij}(a_n, \mathbf{X})$

Her cost increases by
$\beta^t C_n(X_n, a_n, \mathbf{X})$.

Go to the next time step



Strategies and Equilibrium

Stationary Strategy of player n: $\pi^n : \mathcal{S}^N \to \Delta(\mathcal{A})$.
$\pi^n_a(\mathbf{X})$: probability that player n chooses action a under configuration $\mathbf{X}$.
Total cost for n:

$$V_n(\pi^1, \ldots, \pi^N) = \mathbb{E} \sum_{t=0}^{\infty} \beta^i \sum_{a \in \mathcal{A}} C_n(X_n(t), a, \mathbf{X}(t)) \pi^n_a(\mathbf{X}(t)).$$

Solution of the game is often the computation of a Nash equilibrium $(\pi^{1*}, \ldots, \pi^{N*})$ satisfying

$$\forall n, \forall \pi^n, V_n(\pi^{1*}, \ldots, \pi^{n*}, \ldots, \pi^{N*}) \leq V_n(\pi^{1*}, \ldots, \pi^{n}, \ldots, \pi^{N*})$$

Strategies and Equilibrium

Stationary Strategy of player n: $\pi^n : \mathcal{S}^N \to \Delta(\mathcal{A})$.

$\pi^n_a(\mathbf{X})$: probability that player n chooses action a under configuration $\mathbf{X}$.

Total cost for n:

$$V_n(\pi^1, \ldots, \pi^N) = \mathbb{E} \sum_{t=0}^{\infty} \beta^i \sum_{a \in \mathcal{A}} C_n(X_n(t), a, \mathbf{X}(t)) \pi^n_a(\mathbf{X}(t)).$$

Solution of the game is often the computation of a Nash equilibrium $(\pi^{1*}, \ldots, \pi^{N*})$ satisfying

$$\forall n, \forall \pi^n, V_n(\pi^{1*}, \ldots, \pi^{n*}, \ldots, \pi^{N*}) \leq V_n(\pi^{1*}, \ldots, \pi^n, \ldots, \pi^{N*})$$

Theorem (Fink, 1964)

Finite Stochastic games always admit Nash equilibria.

Strategies and Equilibrium

Stationary Strategy of player n: $\pi^n : \mathcal{S}^N \to \Delta(\mathcal{A})$.
$\pi^n_a(\mathbf{X})$: probability that player n chooses action a under configuration $\mathbf{X}$.
Total cost for n:

$$V_n(\pi^1, \dots, \pi^N) = \mathbb{E} \sum_{t=0}^{\infty} \beta^i \sum_{a \in \mathcal{A}} C_n(X_n(t), a, \mathbf{X}(t)) \pi^n_a(\mathbf{X}(t)).$$

Solution of the game is often the computation of a Nash equilibrium $(\pi^{1*}, \dots, \pi^{N*})$ satisfying

$$\forall n, \forall \pi^n, V_n(\pi^{1*}, \dots, \pi^{n*}, \dots, \pi^{N*}) \leq V_n(\pi^{1*}, \dots, \pi^{n}, \dots, \pi^{N*})$$

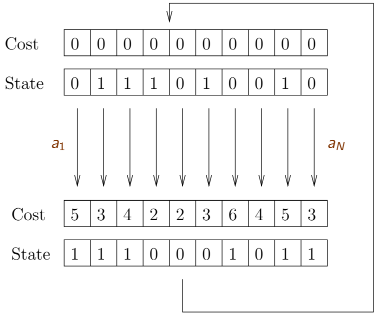
Theorem (Fink, 1964)

Finite Stochastic games always admit Nash equilibria.

In general, hard to get a closed form (even to compute numerically).

Exchangeable Players

Players are exchangeable if P and C_n are insensitive to the labels of the players:
$\forall \sigma \in \mathfrak{S}(N), P(\cdot, \sigma(\mathbf{X})) = P(\cdot, \mathbf{X})$ and $C_n(\cdot, \cdot, \sigma(\mathbf{X})) = C_n(\cdot, \cdot, \mathbf{X})$.

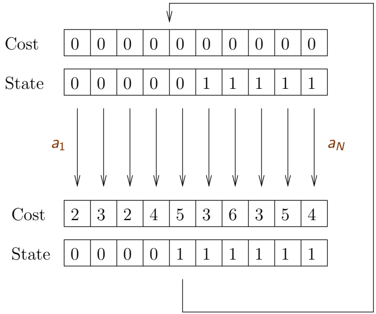


Exchangeable Players

Players are exchangeable if P and C_n are insensitive to the labels of the players:
$\forall \sigma \in \mathfrak{S}(N), P(\cdot, \sigma(\mathbf{X})) = P(\cdot, \mathbf{X})$ and $C_n(\cdot, \cdot, \sigma(\mathbf{X})) = C_n(\cdot, \cdot, \mathbf{X})$.

population distribution is $\mathbf{M} = (M_1, \cdots, M_S) \in \Delta(\mathcal{S})$

$$M_i(t) = \frac{1}{N} \sum_n \delta_{X_n(t)=i}.$$



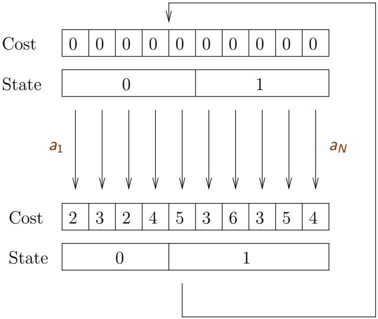
Exchangeable Players

Players are exchangeable if P and C_n are insensitive to the labels of the players:
$\forall \sigma \in \mathfrak{S}(N), P(\cdot, \sigma(\mathbf{X})) = P(\cdot, \mathbf{X})$ and $C_n(\cdot, \cdot, \sigma(\mathbf{X})) = C_n(\cdot, \cdot, \mathbf{X})$.

population distribution is $\mathbf{M} = (M_1, \cdots, M_S) \in \Delta(\mathcal{S})$

$$M_i(t) = \frac{1}{N} \sum_n \delta_{X_n(t)=i}.$$

Dynamics and cost only depend on the population distribution: $P_{ij}(a_n, \mathbf{M})$, $C_n(X_n, a_n, \mathbf{M})$.



Exchangeable Players: Population dynamics

For one player:

$$\mathbb{P}\left(X_n(t+\frac{1}{N})=j \middle| X_n(t)=i, A_n(t)=a, \mathbf{M}(t)=\mathbf{m}\right)=\frac{1}{N}P_{ij}(a,\mathbf{m})$$

Evolution of the population distribution:

$$\mathbb{E}\left(M_j(t+1)|\mathbf{M}(t)\right)= \sum_{i\in\mathcal{S}}\sum_{a\in\mathcal{A}} M_i(t)P_{ij}(a,\mathbf{M}(t))\pi_a(i,\mathbf{M}(t)), \quad (1)$$

Exchangeable Players: Symmetric Strategies

Given a strategy $\pi^0 \in \Pi$ used by player 0 and a strategy $\pi \in \Pi$ used by all the others, $V^N(\pi^0, \pi)$ is the expected discounted payoff of player 0:

$$V^N(\pi^0, \pi) = \\ \mathbb{E}\left[\sum_{t \in \mathcal{T}_N} e^{-\beta t} C_0(X_0(t), A_0(t), \mathbf{M}(t)) \middle| \begin{array}{l} A_0 \text{ has d.b. } \pi^0 \\ A_n \text{ has d.b. } \pi \ (n \neq 0) \end{array} \right].$$

Symmetric Nash Equilibria

A symmetric equilibrium is a strategy π such that no player can gain by deviating from π.

Definition (Symmetric Nash Equilibrium)

For a given set of strategies Π, a strategy $\pi \in \Pi$ is a symmetric Nash equilibrium in Π for the N-player game if, for any strategy $\pi' \in \Pi$,

$$V^N(\pi, \pi) \leq V^N(\pi', \pi).$$

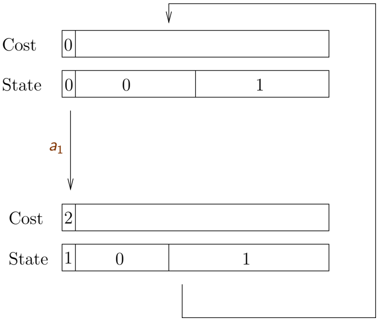
Existence is again guaranteed (Fink, 1964), still not easy to compute.

Let N go to infinity...

Focus on one player (0).
Its influence over the population is $1/N$ and can be neglected, by the rest.

Let N go to infinity...

Focus on one player (0).
Its influence over the population is $1/N$ and can be neglected, by the rest.



Let N go to infinity...(II)

Consequences:

1) players become independent of each other:

$$P(X_0(t) = i_0, X_1(t) = i_1) = P(X_0(t) = i_0)P(X_1(t) = i_1)$$

Let N go to infinity...(II)

Consequences:

1) players become independent of each other:

$$P(X_0(t) = i_0, X_1(t) = i_1) = P(X_0(t) = i_0)P(X_1(t) = i_1)$$

2) The population distribution evolution becomes deterministic:

$$\mathbf{M} \stackrel{\pi}{\to} \mathbf{M}'(:= \mathbf{M} \bullet \pi)$$

with

$$M'_j = \sum_{i \in \mathcal{S}} \sum_{a \in \mathcal{A}} M_i P_{ij}(a, \mathbf{M}) \pi_a(i, \mathbf{M}).$$

Let N go to infinity...(II)

Consequences:
1) players become independent of each other:

$$P(X_0(t) = i_0, X_1(t) = i_1) = P(X_0(t) = i_0)P(X_1(t) = i_1)$$

2) The population distribution evolution becomes deterministic:

$$\mathbf{M} \xrightarrow{\pi} \mathbf{M}'(:= \mathbf{M} \bullet \pi)$$

with

$$M'_j = \sum_{i \in \mathcal{S}} \sum_{a \in \mathcal{A}} M_i P_{ij}(a, \mathbf{M}) \pi_a(i, \mathbf{M}).$$

Kolmogorov Forward Equation

Let N go to infinity...(III)

3) The population strategy is not affected by a single player. Conversely, one player can see the population as a non-strategic environment (the nature) that does not react to its actions.

$$V^{\pi^0,\pi}(x,\mathbf{m}) = \sum_a \pi_a^0(x,\mathbf{m}) C_0(x,a,\mathbf{m}) + \beta \sum_y \sum_a \pi_a^0(x,\mathbf{m}) P_{xy}(a,\mathbf{m}) V^{\pi^0,\pi}(y,\mathbf{m} \bullet \pi)$$

Optimal policy against π (case where π is deterministic):

$$V^{*,\pi}(x,\mathbf{m}) = \min_a \left(C_0(x,a,\mathbf{m}) + \beta \sum_y P_{xy}(a,\mathbf{m}) V^{*,\pi}(y,\mathbf{m} \bullet \pi) \right).$$

Let N go to infinity...(III)

3) The population strategy is not affected by a single player. Conversely, one player can see the population as a non-strategic environment (the nature) that does not react to its actions.

$$V^{\pi^0,\pi}(x,\mathbf{m}) = \sum_a \pi^0_a(x,\mathbf{m}) C_0(x,a,\mathbf{m}) + \beta \sum_y \sum_a \pi^0_a(x,\mathbf{m}) P_{xy}(a,\mathbf{m}) V^{\pi^0,\pi}(y,\mathbf{m}\bullet\pi)$$

Optimal policy against π (case where π is deterministic):

$$V^{*,\pi}(x,\mathbf{m}) = \min_a \left(C_0(x,a,\mathbf{m}) + \beta \sum_y P_{xy}(a,\mathbf{m}) V^{*,\pi}(y,\mathbf{m}\bullet\pi) \right).$$

Bellman Backward Equation

Mean field Equilibrium

The optimal policy in the Bellman equation is the best response to π.

$$BR(\pi) = \arg\min_{\pi^0} V(\pi^0, \pi).$$

Definition (Mean-Field Equilibrium)

A strategy is a (symmetric) mean-field equilibrium if it is a fixed point for the best-response function: $\pi^{MFE} \in BR(\pi^{MFE})$.

Theorem (Existence of equilibrium, Doncel, Gast, G. 2016)

Assume that $P_{ij}(a, \mathbf{m})$ and $C(x, a, \mathbf{m})$ are continuous in $\mathbf{m}$. Then, there always exists a symmetric mean-field equilibrium.

Applying the Kakutani fixed point theorem for infinite dimension spaces to the population distribution (instead of directly to strategies).
Does not require convexity as in Gomes, Mohr, Souza, 2013.

Convergence of continuous policies

Theorem (Convergence, Tembine et Al., 2009)

If $C(i, a, \mathbf{m})$, $P_{ij}(a, \mathbf{m})$ and the policy $\pi_i(\mathbf{m})$ are continuous in $\mathbf{m}$ then the population of the finite game converges to the solution of the Kolmogorov equation.

Furthermore, under such continuity conditions
(i) if π be a mean-field equilibrium, then $\exists N_0$ s.t. $\forall N \geq N_0$, π is a ε-equilibrium of the N player game;
(ii) if $(\pi^N)_N$ is a sequence of Lipschitz-continuous strategies such that π^N is a equilibrium for the N player game, then, any sub-sequence converges weakly to a mean-field equilibrium.

Non-convergence in General

Let us consider a matching game version of the prisoner's dilemma. The state space: $\mathcal{S} = \{A, B\}$ and $\mathcal{A} = \mathcal{S}$. Population distribution is $\mathbf{m} = (m_A, m_B)$.

Cost of a player:

$$C(i, i, \mathbf{m}) = \begin{cases} m_A + 3m_B & \text{if } i = A \\ 2m_B & \text{if } i = B \end{cases}$$

This is the expected cost of a player matched with another player at random and using the cost matrix:

	A	B
A	1, 1	3, 0
B	0, 3	2, 2

(2)

Lemma

There exists a unique mean-field equilibrium π^{∞} that consists in always playing B.

Non-convergence in General (III)

Let us define the following stationary strategy for N players:

$$\pi^N(\mathbf{M}) = \begin{cases} B & \text{if } M_A < 1 \\ A & \text{if } M_A = 1. \end{cases}$$

"play A as long as everyone else is playing A. Play B as soon as another player deviates to B."

Non-convergence in General (III)

Let us define the following stationary strategy for N players:

$$\pi^N(\mathbf{M}) = \begin{cases} B & \text{if } M_A < 1 \\ A & \text{if } M_A = 1. \end{cases}$$

"play A as long as everyone else is playing A. Play B as soon as another player deviates to B."

Lemma

For $\beta < 1$ and N large, π^N is a sub-game perfect equilibrium of the N-player stochastic game.

Non-convergence in General (IV)

Assume that all players, except player 0, play the strategy π^N and let us compute the best response of player 0.
If at time t_0, $M_A < 1$, then the best response of player 0 is to play B.

Non-convergence in General (IV)

Assume that all players, except player 0, play the strategy π^N and let us compute the best response of player 0.
If at time t_0, $M_A < 1$, then the best response of player 0 is to play B.
If $M_A = 1$ then using π, has a cost

$$\frac{1}{N}\sum_{i=0}^{\infty} e^{-\beta i/N} = \int \exp(-\beta t) dt + O(1/N) = 1/\beta + O(1/N).$$

If player 0 chooses action B, all players will also play B after the next step.
This implies that $M_B(t) \approx 1 - \exp(-t)$ and that the player 0 will suffer a cost equal to

$$\int_0^{\infty} (x_A(t) + 2 - 2e^{-t})e^{-\beta t} dt + O(1/N) \geq 2/(\beta(\beta+1)) + O(1/N).$$

Non-convergence in General (IV)

Assume that all players, except player 0, play the strategy π^N and let us compute the best response of player 0.

If at time t_0, $M_A < 1$, then the best response of player 0 is to play B.

If $M_A = 1$ then using π, has a cost

$$\frac{1}{N}\sum_{i=0}^{\infty} e^{-\beta i/N} = \int \exp(-\beta t) dt + O(1/N) = 1/\beta + O(1/N).$$

If player 0 chooses action B, all players will also play B after the next step.

This implies that $M_B(t) \approx 1 - \exp(-t)$ and that the player 0 will suffer a cost equal to

$$\int_0^{\infty} (x_A(t) + 2 - 2e^{-t})e^{-\beta t} dt + O(1/N) \geq 2/(\beta(\beta+1)) + O(1/N).$$

This shows that when $\beta < 1$, player 0 has no incentive to deviate from the strategy π^N so that, π^N is a sug-game perfect equilibrium.

Conclusion

With a finite number of players, it is possible to define many equilibria by using the *"reward and punish"* principle.
These equilibria are linked to the Folk Theorem for repeated games: for each achievable cost, not worse than the Nash equilibrium of the one-shot game, there exists an equilibrium that achieves this cost.

Conclusion

With a finite number of players, it is possible to define many equilibria by using the *"reward and punish"* principle.
These equilibria are linked to the Folk Theorem for repeated games: for each achievable cost, not worse than the Nash equilibrium of the one-shot game, there exists an equilibrium that achieves this cost.

When the number of players is infinite, the deviation of a single player is not visible by the population, the equilibria based on the "tit for tat" principle do not scale at the mean-field limit.

Conclusion

With a finite number of players, it is possible to define many equilibria by using the *"reward and punish"* principle.
These equilibria are linked to the Folk Theorem for repeated games: for each achievable cost, not worse than the Nash equilibrium of the one-shot game, there exists an equilibrium that achieves this cost.

When the number of players is infinite, the deviation of a single player is not visible by the population, the equilibria based on the "tit for tat" principle do not scale at the mean-field limit.

This is all the more damaging because these equilibria have very good social costs: mean-field games fail to describe the best equilibria.
Or, in other words, mean field games cannot fight against free riders.

One example: Epidemic Model with Vaccinations

- A player encounters other players with rate γ.
- If the first player is susceptible and the second is infected, the first one becomes infected.
- An infected player recovers at rate ρ.
- A susceptible player can get vaccinated with strategy π, where $\pi(t) \in [0, \tau]$.
- Once an player is vaccinated or recovered, her state does not change.

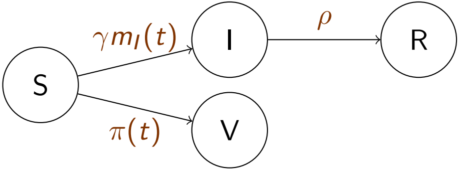


Figure : S (susceptible), I (infected), R (recovered) and V (vaccinated).

Mean Field Model

In continuous time, the systems converges to a system of differential equations.

$$\begin{cases} \dot{m}_S^\pi(t) = -\gamma m_S^\pi(t) m_I^\pi(t) - \pi(t) m_S^\pi(t) \\ \dot{m}_I^\pi(t) = \gamma m_S^\pi(t) m_I^\pi(t) - \rho m_I^\pi(t) \\ \dot{m}_R^\pi(t) = \rho m_I^\pi(t) \\ \dot{m}_V^\pi(t) = \pi(t) m_S^\pi(t) \end{cases}$$

Single player Cost

Player 0 using policy π^0 has an evolution given by:

$$\begin{cases} \dot{x}_S^{\pi^0,\pi}(t) = -\gamma x_S^{\pi^0,\pi}(t) m_I^{\pi}(t) - \pi(t) x_S^{\pi^0,\pi}(t) \\ \dot{x}_I^{\pi^0,\pi}(t) = \gamma x_S^{\pi^0,\pi}(t) m_I^{\pi}(t) - \rho x_I^{\pi^0,\pi}(t) \\ \dot{x}_R^{\pi^0,\pi}(t) = \rho x_I^{\pi^0,\pi}(t) \\ \dot{x}_V^{\pi^0,\pi}(t) = \pi(t) x_S^{\pi^0,\pi}(t) \end{cases}$$

Using the foregoing notations, the expected individual cost of Player 0 is defined as follows:

$$V(\pi^0, \pi) = \int_0^T \left(c_V \pi^0(t) x_S^{\pi^0,\pi}(t) + c_I x_I^{\pi^0,\pi}(t) \right) dt,$$

where c_V is the vaccination cost and c_I is the unit time cost of being infected.

Bellman Best Response Equation of One Player

$$J_S(t) = \inf_{\pi^0(t)\in[0,\tau]} \left[G(t) + \frac{\pi^0(t)}{\mu} \left(c_V - J_S(t+\frac{1}{\mu}) \right) \right], \quad (3)$$

$$J_I(t) = \frac{c_I}{\mu} + J_I(t+\frac{1}{\mu}) \left(1 - \frac{\rho}{\mu} \right), \quad (4)$$

where $J_S(T) = J_I(T) = 0$ and

$$G(t) = \left(1 - \frac{\gamma I(t)}{\mu} \right) J_S(t+\frac{1}{\mu}) + \frac{\gamma I(t)}{\mu} J_I(t+\frac{1}{\mu}). \quad (5)$$

Nash Equilibria

Lemma

For any population strategy π, there exists a best-response π^0 that is a threshold strategy.

Theorem

There exists a unique symmetric mean-field equilibrium that is pure and of threshold type.

Centralized Optimal Policy

We denote by $C(\pi)$ the cost incurred in the system by the population vaccination strategy π, i.e.,

$$C(\pi) = \int_0^T (c_I m_I(t) + c_V \pi(t) m_S(t)) \, dt.$$

The global optimum of the problem is the population strategy that minimizes the total cost and let

$$\pi^{opt} \in \arg\min_{\pi} C(\pi).$$

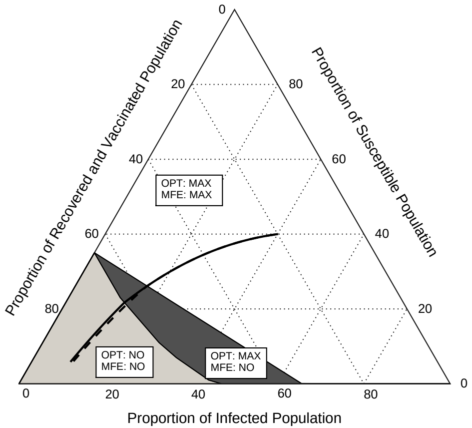
As for the case of mean-field equilibrium, a global optimum is a threshold strategy.

Proposition

The strategy that minimizes the total cost is a threshold strategy.

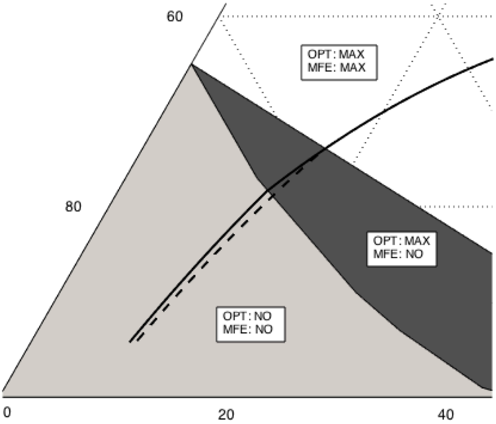
Numerical Comparisons

Using $\rho = 36.5$, $\gamma = 73$, $\tau = 10$, $T = 0.3$, $c_I = 36.5$ and $c_V = 0.5$.

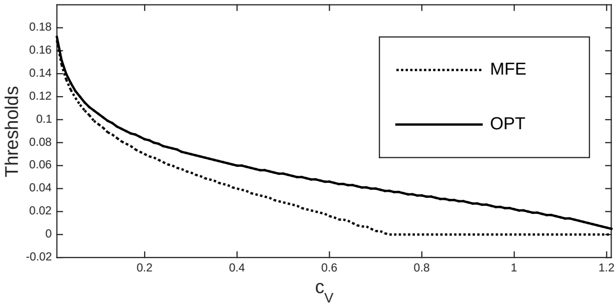


Population dynamics under the equilibrium strategy (dashed line) and the global optimum strategy (solid line).

Numerical Comparisons



Numerical Comparisons



Thresholds of the MFE and of the global optimum when $c_V \in [0.01, 1.21]$.

If the vaccination decisions are let to individuals, then vaccination should be subsidized, by removing a cost p so that both thresholds coincide, *i.e.*,

$$t^{eq}(c_V - p) = t^{opt}(c_V).$$

(Here, $p = 0.35$).