Modeling Opponents

in

Adversarial Risk Analysis

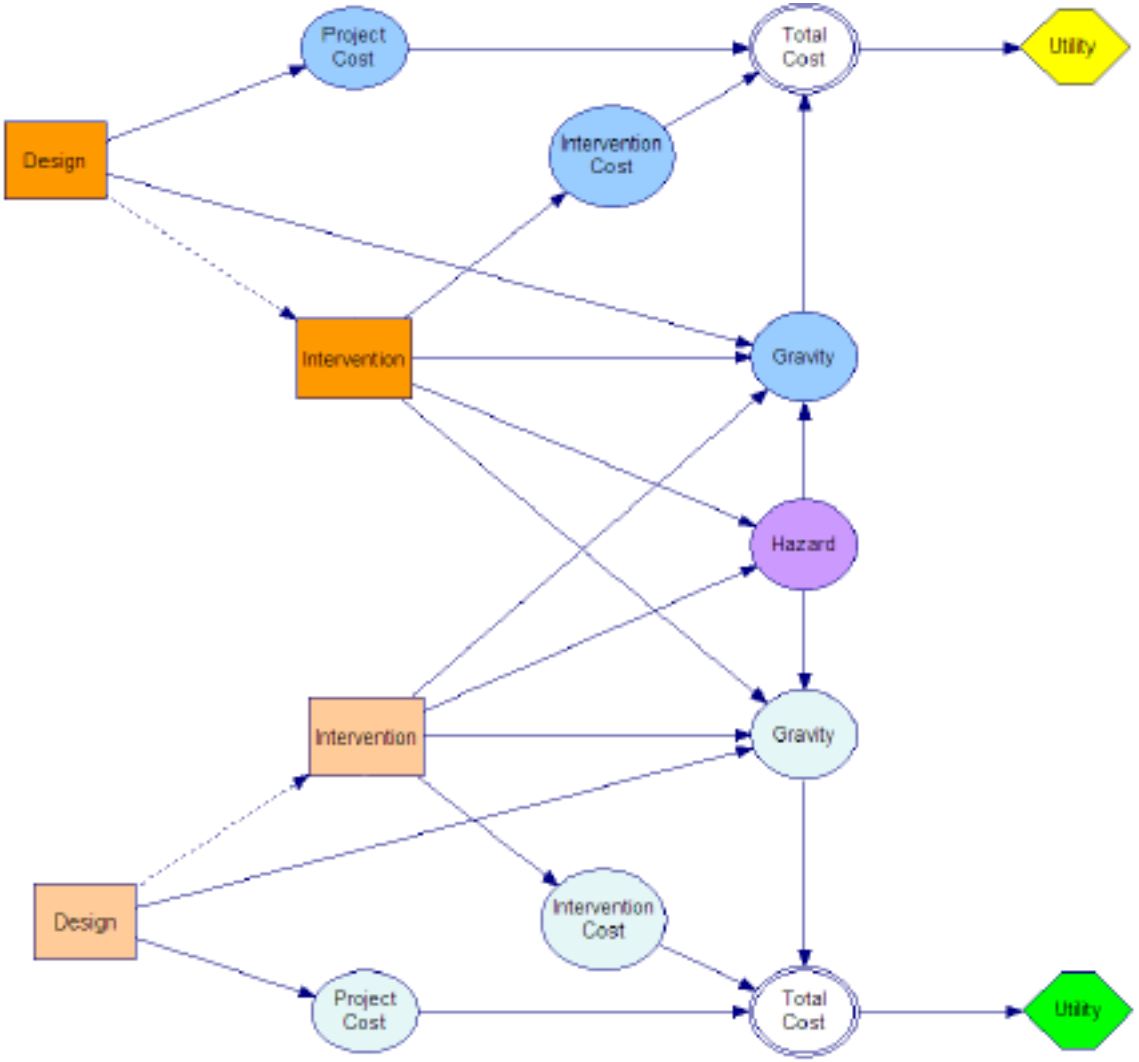
David Ríos Insua, R. Academy
David Banks, Duke U.
Jesus Rios, IBM Research YH

ADT 2013, Brussels November

Outline

- ARA
- The basic problem
- A few opponent models
  - Non strategic
  - Nasheq
  - Level-k
  - Mirroreq
  - Prospect
- Reconciling and Learning
- Discussion

Adversarial risk analysis



ARA

- RA enhanced to include adversaries ready to increase our risks

- S-11, M-11 lead to large security investments globally
- Many modelling efforts to efficiently allocate such resources
- Parnell et al (2008) NAS review
  - Standard reliability/risk approaches not take into acocunt intentionality
  - Game theoretic approaches. Common knowledge assumption…
  - Decision analytic approaches. Forecasting the adversary action…
- Merrick, Parnell (2011) review approaches commenting favourably on Adversarial Risk Analysis

- SECONOMICS FP7 project

ARA

- A framework to manage risks from actions of intelligent adversaries (DRI, Rios, Banks, JASA 2009)

- One-sided prescriptive support
  - Use a SEU model
  - Treat the adversary's decision as uncertainties

- Method to predict adversary's actions
  - We assume the adversary is a *expected utility maximizer*
    - Model his decision problem
    - Assess his probabilities and utilities
    - Find his action of maximum expected utility
  - But other *descriptive* models are possible….

- Uncertainty in the Attacker's decision stems from
  - *our* uncertainty about his probabilities and utilities
  - but this leads to a hierarchy of nested decision problems

ARA

- ARA applications to counterterrorism models (Rios, DRI, 2012)
  - Sequential Defend-Attack, Simultaneous Defend-Attack, Sequential Defend-Attack-Defend, Sequential Defend-Attack with private information

- Somali pirates (Sevillano, Rios, DRI, 2012)
- Routing games (anti IED war) (Wang, Banks, 2011)
- Airport security (Cano, DRI, Tedeschi, Turhan, 2013). SECONOMICS
- Metro security (Cano, DRI, Pellot, 2013). SECONOMICS
- Urban security (Gil, DRI, Rios, 2013).

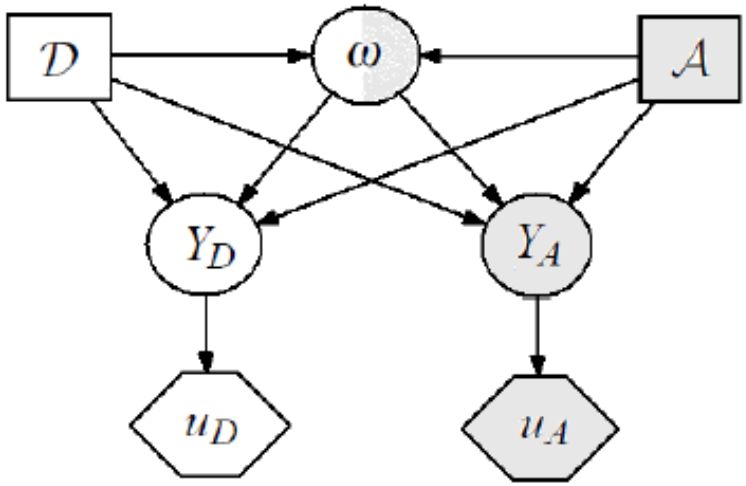
- Borel games (Banks, Petralia, Wang, 2011)
- Auctions (DRI, Rios, Banks, 2009; Banks, Rios, DRI, 2014)
- Social robotics (Razuri, Esteban, DRI, 2012)

- Kadane, Larkey (1982), Raiffa (1982), Lippman, McCardle (2012), Rothkopf (2007)
- Stahl and Wilson (1994,1995)
- Wolpert and Lee (2012)
- Rotschild, MacLay, Guikema (2012)

- **But other *descriptive* models are possible**….

Basic Problem

$\{d_1, \ldots, d_m\}$



$\{a_1, \ldots, a_n\}$

$$\psi_D(d,a) = \int u_D(a,d,\omega) p_D(\omega|a,d) d\omega$$

$$\psi_A(d,a) = \int u_A(a,d,\omega) p_A(\omega|a,d) d\omega$$

	a
d	$(\psi_D(d,a), \psi_A(d,a))$

$$\max_d \psi_D(d) = \sum_{i=1}^{n} \psi_D(d,a_i) p_D(a_i) = \sum_{i=1}^{n} \left[\int u_D(a_i,d,\omega) p_D(\omega|a_i,d) d\omega \right] p_D(a_i)$$

$p_D(a)$

Non strategic opponent. I

- A lacks memory. Dirichlet-multinomial model

$$(p_1, \ldots, p_n) \sim \mathcal{D}(\alpha_1, ..., \alpha_n)$$

$$(p_1, \ldots, p_n)|\text{data} \sim \mathcal{D}(\alpha_1 + h_1, ..., \alpha_n + h_n)$$

$$p_D^{NS}(a_i) = E(p_i|\text{data}) = \frac{\alpha_i + h_i}{\sum_{j=1}^{n}(\alpha_j + h_j)}, i = 1, \ldots, n.$$

$$\max_d \sum_{i=1}^{n} \psi_D(d, a_i) p_D^{NS}(a_i)$$

Non strategic opponent. II

- A remembers his last attack, her last defense and the results. Matrix-beta prior model

$$(p_1, \ldots, p_n)|a_i, d_j, \omega \sim \mathcal{D}(\alpha_1^{ij\omega}, ..., \alpha_n^{ij\omega})$$

$$(p_1, \ldots, p_n)|a_i, d_j, \omega, \text{data} \sim \mathcal{D}(\alpha_1^{ij\omega} + n_1^{ij\omega}, ..., \alpha_n^{ij\omega} + n_n^{ij\omega})$$

- To control size growth, mixture model

$$p_D(a|a_i, d_j, \omega) = w_1 p_D(a|a_i) + w_2 p_D(a|d_j) + w_3 p_D(a|\omega).$$

Inference through a Gibbs sampler

Ficticious play

Nasheq opponent

- Opponent computes Nash equilibria

$$(U_A, P_A) \qquad (U_D, P_D)$$

- For each $\theta \in \Theta$

$$\Psi_D^\theta(d,a) = \int U_D^\theta(a,d,\omega) P_D^\theta(\omega|a,d) d\omega. \qquad (\psi_D^\theta(d,a), \psi_A^\theta(d,a))$$

$$(d^N(\theta), a^N(\theta)) \qquad p_D^N(a) = E_{\mathcal{P}}(a^N(\theta)).$$

$$\max_d \sum_a \psi_D(d,a) p_D^N(a)$$

Level-k thinking opponent I

- D needs to solve

$$d^* = \arg\max_d \left[\sum_a \psi_D(d,a) p_D(a) \right]$$

- For this, she thinks about A's problem

$$\begin{aligned} a^* &= \arg\max_a \left[\sum_d \psi_D(d,a) p_A(d) \right] \\ &= \arg\max_a \left[\sum_d \int u_A(d,a,\omega) p_A(\omega|a,d) d\omega \right] p_A(d) \end{aligned}$$

- She does not know $(u_A, p_A(.|.), p_A)$
- Models uncertainty through $(U_A, P_A(.|.), P_A)$

$$A|D \sim \arg\max_a \sum_d \left[\int U_A(d,a,\omega) P_A(\omega|a,d) d\omega \right] P_A(d) \qquad p_D(a) = p_{A|D}(a)$$

Level-k thinking opponent II

$$(U_A, P_A(\cdot \mid \cdot), P_A)$$

$$D|A^1 \sim \arg\max_d \sum_a \left[\int U_D(d,a,\omega) P_D(\omega|a,d) d\omega \right] P_D(a),$$

Repeat from $i = 1$

Find $P_{D^{i-1}}(A^i)$ by solving

$$A^i \mid D^i \sim \arg\max_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\int U_A^i(a,d,\omega) P_A^i(\omega \mid a,d) d\omega \right] P_A^i(D^i = d)$$

with $(U_A^i, P_A^i(\cdot \mid \cdot), P_A^i) \sim F^i$

Find $P_A^i(D^i)$ by solving

$$D^i \mid A^{i+1} \sim \arg\max_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\int U_D^i(a,d,\omega) P_D^i(\omega \mid a,d) d\omega \right] P_D^i(A^{i+1} = a)$$

with $(U_D^i, P_D^i(\cdot \mid \cdots), P_D^i) \sim G^i$

$i = i + 1$

MacLay, Rothschild, Guikema (2012) Rios, DRI (2012)

Mirroreq opponent

$(U_A, P_A(\cdot \mid \cdot), P_A(\cdot))$ $(U_D, P_D(\cdot \mid \cdot), P_D(\cdot))$

$$
\begin{aligned}
d^M(\theta) &= \arg\max_d \sum_d \left[\int U_D^\theta(d,a,\omega) P_D^\theta(\omega \mid a,d) d\omega \right] p_D^M(a) \\
p_D^M &= E_{\mathcal{P}} \left(\arg\max_a \sum_d \left[\int U_A^\theta(d,a,\omega) P_A^\theta(\omega \mid a,d) d\omega \right] P_A^{\theta M}(d) \right)
\end{aligned}
$$

$$
\max_d \sum_a \psi_D(d,a) p_D^M(a)
$$

Prospect opponent

- EU model OK for D (as giving prescriptive advice)
- EU model OK for A???
- Terrorist psychology and logistics suggest optimising terrorists (cutthroat capitalism)

$$\arg\max_a \left[\sum_d \int v_A(d,a,\omega) w_A^1(p_A(\omega|a,d)) d\omega \right] w_A^2(p_A(d))$$

$$A|D = \arg\max_a \left[\sum_d \int V_A(d,a,\omega) W_A^1 P_A(\omega \,|\, a,d)) d\omega \right] W_A^2(P_A(d))$$

Reconciling and learning about opponent model

- Use a mixture of opponent models

$$p_D(a) = \sum_{i=1}^{k} q_i p_D^i(a)$$

- Model averaging to optimize

$$\max_d \sum_a \psi_D(d,a) \left(\sum_{i=1}^{k} q_i p_D^i(a) \right) = \max_d \sum_{i=1}^{k} q_i \left[\sum_a \psi_D(d,a) p_D^i(a) \right]$$

- Model uncertainty to learn about weights.

Discussion

General extensions

- Non-discrete. Easy
- Non-simultaneous. Not so easy, if general structure.
- Several adversaries. Not coordinated, coordinated.

Technical issues

- Fixed points for mirroreq
- Value of climbing up one level in k-thinking