Non-Bayesian Persuasion

Geoffroy de Clippel, Xu Zhang('22 JPE)

Renjie Zhong

PPE

RUC

February 2023

Outline

Introduction

(Bayesian) Persuasion: Focus

- (optimal) persuasion:
  1. Process: bayesian updating and bayesian plausibility (martingale property/consistency/committment)
  2. Characteristics: cardinality of signals, posterior distribution...
  3. Approach: concavification (AM,95; KG2011) and BCE/Myersionian/RP (KG,11;BM,16,19)
- welfare:
  1. when persuasion is beneficial to the sender
  2. when it is detrimental to the receiver

Persuasion: (KG,2011)' s Insights

- choose distributions of distributions (posteriors): a geometric approach (concavification)
- when does persusion benefits the sender and receivers:
  1. Sender: action under no info is not preferred by sender and is constant in some neighborhood of beliefs around the prior
  2. Receiver: never detrimental
- (BM,16,19) uses a **Myersionian Approach** to reexpress this problem, thus conveniently processing high-dimensional problems
- Robustness?

Persuasion: Anomalies from Real Life

- updating rules:
  1. 油盐不进 (a stubborn receiver): 范增说项羽, 晁盖命丧曾头市
  2. 耳根子软/轻信于人 (a gullible receiver): 屈原投江
  3. different interpretations:
     1. motivated updating (endogeneous choice): $\mu_s^R(\omega;\mu_0,\pi) = \underset{\hat{\nu}\in\Gamma(\nu)}{\operatorname{argmax}} \mathcal{U}(\hat{\nu},\nu,\nu^*)$,
     2. conservative updating: $\mu_s^R(\omega;\mu_0,\pi) = (1-\chi)\mu_0 + \chi\nu$
     3. $\alpha-\beta$ updating: $\mu_s^R(\omega;\mu_0,\pi) = \frac{\pi(s|\omega)^\beta \mu_0(\omega)^\alpha}{\sum_{\omega'\in\Omega}\pi(s|\omega')^\beta \mu_0(\omega')^\alpha}$
- welfare:
  1. 苏秦张仪周游列国
  2. 群英会蒋干中计
  3. 严嵩哭求夏言

Persuasion: Reinvestigation

- an extended set of updating rules: definition? property?
- robustness of approaches: concavification and revealation principle
- welfare:
    1. when persuasion is beneficial to the sender
    2. when it is detrimental to the receiver
    3. an order relation of updating rules?

Non-Bayesian Persuasion

- Restriction to so-called Systematical Distortion Updating rules: $\mu^R = D_{\mu_0}(\mu^B)$
  1. independent of the label used to describe that realization (neutrality)
  2. independent of irrelevant signal realizations
  3. homogeneity of degree zero in the likelihood of getting that signal realization as a function of the different states
- the revealation principle often fails (with non-systematical one and some systematical rules) while the concavification holds robust with systematical one

  concavification: transformation links μ^S and μ^R

  revealtion principle: bayesian nature - convexity - affine property
- Order: no two systematically distorted rules can be unambiguously compared when permitting all payoff structures

Non-Bayesian Persuasion

- Probability: unambiguous (considering two extreme situations: stubborn/gullible)
- Benefits:

  Sender: (KG,2011) holds robust and overinference bias always benefits the sender

  Receiver: detrimental persuasion occurs only when conflicting interests exist

  with mixed interests, ambiguous

Distortion Rules

Setup

- state: $\omega \in \Omega$, VNM: $u(a,\omega)$, $v(a,\omega)$, prior: μ_0
- information structure: $\{\pi(\cdot|\omega)\}_{\omega\in\Omega}$ over $s \in S$
- bayesian updating: $\mu^S = \mu_s^B(\omega;\mu_0,\pi) = \frac{\pi(s|\omega)\mu_0(\omega)}{\sum_{\omega'\in\Omega}\pi(s|\omega')\mu_0(\omega')}$
- signal induces a posterior pair: $\tau \in \Delta(\Delta(\Omega) \times (\Delta(\Omega))$

  $Pr((\nu,\nu')) = \sum_{s\in S(\nu,\nu')}\sum_{\omega}\pi(s \mid \omega)\mu_0(\omega), \nu = \mu^B, \nu' = \mu^R$

  $T(\mu_0,\mu_R)$; the set of all such distributions obtained by varying π

Systematically Distorted Updating

- distortion function: $D_{\mu_0} : \Delta(\Omega) \to \Delta(\Omega)$ such that for all μ_0
  $\mu^R(\cdot;\mu_0,\pi) = D_{\mu_0}(\mu^S(\cdot;\mu_0,\pi))$ for all signals π and all signal realizations s
  1. independent of the signal /neutrality
  2. independence of irrelevant signal realizations

Systematically Distorted Updating

The updating rule μ^R systematically distorts updated beliefs if and only if, given any full-support prior μ_0 , $\mu^R_s(\cdot;\mu_0,\pi) = \mu^R_{\hat{s}}(\cdot;\mu_0,\hat{\pi})$ for all signal realization pairs (π,s) and $(\hat{\pi},\hat{s})$ such that the likelihood ratio $\frac{\hat{\pi}(\hat{s},\omega)}{\pi(s,\omega)}$ is constant as a function of ω

- The likelihood of getting that signal realization as a function of the different states matters
- The updated belief should remain unchanged when rescaling those probabilities by a common factor, a property of homogeneity of degree zero.

Examples

1. motivated updating (endogeneous choice): $D_{\mu_0}^{MU}(\nu) = \underset{\hat{\nu}\in\Gamma(\nu)}{\operatorname{argmax}} \mathcal{U}\left(\hat{\nu}, \nu, \nu^*\right)$
2. conservative updating: $D_{\mu_0}^{CB} = (1-\chi)\mu_0 + \chi\nu$

   affine updating: $D_{\mu_0}^{\chi,\nu^*} = (1-\chi)\nu^* + \chi\nu$
3. $\alpha - \beta$ updating: $\mu_s^R\left(\omega;\mu_0,\pi\right) = \frac{\pi(s|\omega)^\beta \mu_0(\omega)^\alpha}{\sum_{\omega'\in\Omega}\pi(s|\omega')^\beta \mu_0(\omega')^\alpha}$

   α: base rate negalect/overweighting prior, β: over/underinference

   $D_{\mu_0}^{\alpha,\beta}(\nu) = \frac{\nu^\beta \mu_0^{\alpha-\beta}}{\sum_{\omega'\in\Omega}\nu(\omega')^\beta \mu_0(\omega')^{\alpha-\beta}}$

Broader Examples: (AC,2016)‘s insights

1. different priors: $\mu_s^R(\cdot;\mu_0,\pi)=\mu_s^B(\cdot;\mu_0^R,\pi)$
   $$D_{\mu_0}^{NCP}(\nu)=\frac{\nu\left(\mu_0^R/\mu_0\right)}{\sum_{\omega'\in\Omega}\nu(\omega')\left(\mu_0^R(\omega')/\mu_0(\omega')\right)}$$
2. probability weighting: $W(\mu_s^B(\cdot;\mu_0,\pi)), W:\Delta(\Omega)\to\Delta(\Omega)$

Counterexamples

1. No learning without full disclosure:

$$\mu_s^R(\omega;\mu_0,\pi)=\begin{cases} 1 & \text{if } \pi(s\mid\omega)>0 \text{ and for all } s', \exists!\omega' \text{ such that } \pi\left(s'\mid\omega'\right)>0, \\ 0 & \text{if } \pi(s\mid\omega)=0 \text{ and for all } s', \exists!\omega' \text{ such that } \pi\left(s'\mid\omega'\right)>0 \\ \mu_0(\omega) & \text{otherwise.} \end{cases}$$

2. Normalized transformation: $\mu_s^f(\omega;\mu_0,\pi)=\frac{f(\pi(s|\omega),\mu_0(\omega))}{\sum_{\omega'\in\Omega} f(\pi(s|\omega'),\mu_0(\omega'))}$
3. Information aggregation mistakes: $\mu_s^{AVG}(\cdot;\mu_0,\pi)=\sum_{k=1}^{K}\frac{1}{K}\mu_{s_k}^B(\cdot;\mu_0,\pi_k)$
4. Correlation neglect: $\mu_s^{CN}(\cdot;\mu_0,\pi)=\mu_s^B\left(\cdot;\mu_0,\prod_{k=1}^{K}\pi_k\right)$
5. Rational Inattention

all that matters is how states correlate with signal realizations s, and the nature of those realizations does not matter.

Concavification and Revealation Principle

Optimal Persuasion

- Optimal action for player with ν': $\hat{a}\left(\nu'\right) \in argmax_{a\in A} E_{\nu' u(a,\omega)}$
- Optimal persuasion: $V\left(\mu_0, \mu^R\right) = \sup_{\tau\in T(\mu_0,\mu^n)} E_\tau \hat{v} = \sup_{\tau\in T(\mu_0,\mu^n)} \sum_{(\nu,\nu')\in \text{supp}(\tau)} \tau\left(\nu,\nu'\right)\hat{v}\left(\nu,\nu'\right)$
  where $\hat{v}\left(\nu,\nu'\right) = \sum_\omega \nu(\omega) v\left(\hat{a}\left(\nu'\right),\omega\right)$
- Remark(state-independent sender's utility): $\hat{v}\left(\nu,\nu'\right) = v(\hat{a}(\nu'))$
  1. only $\tau^R \in T^R\left(\mu_0,\mu^R\right)$, $\tau^R \in \Delta(\Delta(\Omega))$ matters and $\tau^R = \sum_{(\nu,\nu')\in \text{supp}(\tau)} \left(\tau\left(\nu,\nu'\right)\right)$
  2. $V\left(\mu_0,\mu^R\right) = \sup_{\tau^R\in T^R\left(\mu_0,\mu^R\right)} E_{\tau^R}\hat{v} = \sup_{\tau^R\in T^R\left(\mu_0,\mu^R\right)} \sum_{\nu'\in supp\left(\tau^R\right)} \tau^R\left(\nu'\right)\hat{v}\left(\nu'\right)$

Concavification

- Concavification: $[\mathrm{CAV}(f)](\mu) = \sup\{z \mid (\mu, z) \in \mathrm{co}(f)\}$

Beneficial Persuasion

The sender benefits from persuasion iff $[\mathrm{CAV}(\check{v})](\mu_0) > \hat{v}(\mu_0, \mu_0)$, $\check{v} = \hat{v}(\nu, D_{\mu_0}(\nu))$

- $\hat{v}(\mu_0, \mu_0) = \sum_\omega \mu_0(\omega) v(\hat{a})(\mu_0, \omega) \neq \sum_\omega \mu_0(\omega) v(\hat{a})(D_{\mu_0}(\mu_0), \omega)$

RP: not systematically distorted updating rules

RP: not systematically distorted updating rules

The revelation principle fails if μ^R does not systematically distort updated beliefs and μ^R satisfies:

1. $\mu_s^R(\cdot;\mu_0,\pi)$ is a continuous function of the vector $\{\pi(\cdot|\omega)\}_{\omega\in\Omega} \in R_+^{|\Omega|}\backslash\{0\}$
2. the receiver is certain that state ω occurs after a realization s of some experiment if and only if s occurs with strictly positive probability only in state ω

- updated beliefs associated with a signal realization s depends on only the state-dependent probability of s in the different states under the experiment
- violate neutrality + independence of irrelevant signal realizations

RP: systematically distorted updating rules

- $\Omega = \{\omega_1, \omega_2, \omega_3\}$, μ_0. $A = \{a_1, a_2\}$
- $\mu_s^R(\omega; \mu_0, \pi) = \frac{\pi(s|\omega)^2 \mu_0(\omega)}{\sum_{\omega' \in \Omega} \pi(s|\omega')^2 \mu_0(\omega')}$, $D_{\mu_0}^{1,2}(\nu) = \frac{\nu^2 \mu_0^{-1}}{\sum_{\omega' \in \Omega} \nu(\omega')^2 \mu_0(\omega')^{-1}}$

	Signal			Receiver's Utility u	
	s_1	s_2	s_3	a_1	a_2
ω_1	1	0	0	ρ	0
ω_2	0	1	0	ρ	0
ω_3	φ	φ	$1-2\varphi$	0	1

- $0 < \phi < 0.5$ and $\phi^2 < \rho < 2\phi^2$, the receiver strictly prefers a_1 upon realizations s_1 and s_2 and strictly prefers a_2 upon s_3
- $S^{a_1} = \{s_1, s_2\}$ fails to recommend a_1

RP: systematically distorted updating rules

1. the optimal signal does require three realizations
2. the optimal signal involves only two realizations but cannot simply recommend an action that the receiver will follow
3. the revelation principle fails, but the optimal signal gives an incentive-compatible action recommendation

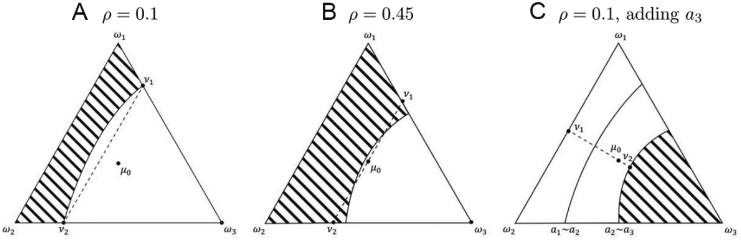


FIG. 1.—Illustration of sender's distorted indirect utility function $\hat{\nu}$. $\hat{\nu} = 1$ in the diagonally striped area (boundaries included); $\hat{\nu} = 0$ elsewhere.

RP: systematically distorted updating rules

RP and Convexity

Given Ω, if the distortion function D_{μ_0} satisfies that for any ν_1 and $\nu_2 \in \Delta\Omega$ and any $\lambda \in [0, 1]$, there exists $\gamma \in [0, 1]$ such that

$$D_{\mu_0}(\lambda\nu_1 + (1-\lambda)\nu_2) = \gamma D_{\mu_0}(\nu_1) + (1-\gamma)D_{\mu_0}(\nu_2) \quad (1)$$

Then the revelation principle holds for all persuasion problems with prior μ_0

- any Bayesian posterior induced by a recommendation a is a convex combination of the Bayesian posteriors induced by the original signal realizations in S^a
- convexity makes a remain optimal
- affine distortion function satisfies this condition and $\alpha - \beta$ with $\beta = 1$

RP: pther discussions

RP and Convexity

Given $|\Omega| \geq 3$, if there exist two beliefs ν_1 and $\nu_2 \in \Delta\Omega$ and $0 < \lambda < 1$ such that $D_{\mu_0}(\lambda\nu_1 + (1-\lambda)\nu_2)$ is not collinear with $D_{\mu_0}(\nu_1)$ and $D_{\mu_0}(\nu_2)$, then the revelation principle fails.

If one-to-one distortion function then RP holds iff D_{μ_0} is projective transformation

- $|A| < |S| \leq |\Omega|$ if RP fails, $|S| \leq |A|$ if RP holds

Order of Updating Rules

Prefernce

In persuasion problem $(\Omega, \mu_0, A, (u, v), \mu_R)$

- unambiguously prefer μ^R over $\hat{\mu}^R$: $\hat{v}(\pi', \mu_R) \geq \sup_\pi \hat{v}(\pi, \hat{\mu}^R)$ for all action sets A, all utility functions (u, v)
- μ^R is easier to persuade than $\hat{\mu}^R$: $T\left(\mu_0, \hat{\mu}^R\right) \subseteq T\left(\mu_0, \mu^R\right)$
- feasible belief: $\nu \neq \mu_0, \nu \in T\left(\mu_0, \mu^R\right)$

Intuitive Ranking

1. The sender unambiguously prefers μ^R over $\hat{\mu}^R$ if μ^R is easier to persuade than $\hat{\mu}^R$
2. if the sender unambiguously prefers μ^R over $\hat{\mu}^R$, then one cannot find a posterior that is feasible for $\hat{\mu}^R$ but not for μ^R

Incomparablity between Systmatically Distorted Rules

Incomparablity

μ^R and $\hat{\mu}^R$ corresponds to two one-to-one systematically distorted updating rules, then neither $\mu^R \succcurlyeq \hat{\mu}^R$ and $\hat{\mu}^R \succcurlyeq \mu^R$

- what matters for unambiguous preference comparisons are distributions over posterior pairs
- {过分易打动/顽固的人}×{观点一致/不一致的场景}

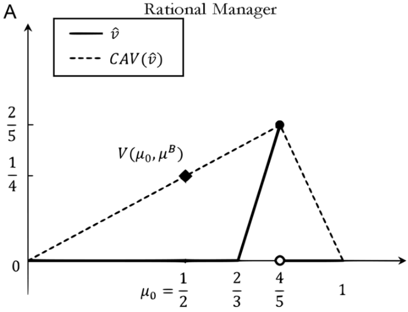
Example: Bayesian and Conservative Bayesian

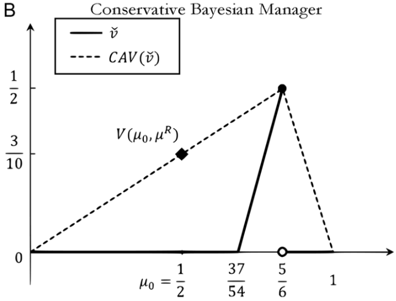
PLAYERS' PAYOFFS

	Abe	Bob	No One
	A. Manager (Receiver)		
ω_1	1	2	0
ω_2	-2	-6	0
	B. Abe (Sender)		
ω_1	1	0	0
ω_2	-2	0	0

- conservative updating: $D^{CB}_{\mu_0}(\nu) = \frac{1}{10}\mu_0 + \frac{9}{10}\nu$

Example: Bayesian and Conservative Bayesian





Smaller Classes of Persuasion Problems

1. Sender's Utility Is State Independent
2. Purely Opposed Interests: running no experiment at all is optimal
3. Common Interests: all rules that correctly update beliefs in fully revealing experiments

$$\begin{cases} D_{\mu_0}(\nu) = \chi(\nu)\mu_0 + (1-\chi(\nu))\nu \\ \hat{D}_{\mu_0}(\nu) = \hat{\chi}(\nu)\mu_0 + (1-\hat{\chi}(\nu))\nu \end{cases}$$

$$D_{\mu_0}(\nu) = \hat{\delta}\hat{D}_{\mu_0}(\nu) - \delta\mu_0$$

$$T(\mu_0, \mu^{CB\chi}) \text{ strictly decreasing in } \chi$$

4. Getting the Receiver to Switch Action

Welfare of Senders and Receivers

When Does the Sender Benefit from Persuasion?

- unambiguously prefer μ^R over $\hat{\mu}^R$: $\hat{\mu}^R$ benefits sender $\Rightarrow$ μ^R benefits sender
- discrete preference for receivers: if there is an $\epsilon > 0$ such that $\forall a \neq \hat{a}(\nu'), E_{\nu'} u(\hat{a}(\nu'), \omega) > E_{\nu'} u(a, \omega) + \varepsilon$
- the desire to share information for senders: $\hat{v}(\nu, D_{\mu_0(\nu)}) \geq \hat{v}(\nu, \mu_0)$
- regular: D_{μ_0} is continuous and μ_0

Beneficial Persuasion

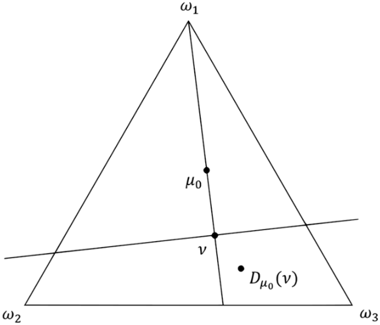
Fix μ_0 and regular systematically distorted function

1. if no information the sender would share at μ_0, then the sender does not benefit from persuasion
2. if there exists information the sender would share and the receiver's preference is discrete at μ_0 (which is generically true when A is finite), then the sender benefits from persuasion.

On the Possibility of Detrimental Persuasion

An Intuitive Outcome

Suppose that there exist $\omega \in \Omega$ and $\mu_0, \nu \in \Delta(\Omega)$ such that D_{μ_0} is continuous,μ_0 is a strict convex combination of μ_0 and δ_ω, and $(D_{\mu_0}(\nu) - \nu)(\delta_\omega - \nu) < 0$. Then optimal persuasion is harmful to the receiver in some switch action problem.



On the Possibility of Detrimental Persuasion

An Intuitive Outcome

Optimal persuasion is never detrimental to the receiver in case of conservative Bayesianism.

Extensions

- optimal persuasion is possible with stochastic distortion rules
- robust persuasion holds with corresponds

Conclusion

Conclusions

- Conclusions:
  1. Concavification: natural extension
  2. Revealation Principle: convexity may not be robust
  3. Ranking: intuitive requirements and incomparability
  4. Beneficial persuasion for senders: desire and possibility
  5. Detrimental persuasion for receivers: very unambiguos
- a good way to differentiate mechanism design with information design
- the computationality of picking up optimal posteriors is hard (especially with high dimensions)
- group persuasion: who‘s who wanted in the persuasion?