Robust Bayesian Truth Serum

Presentation by Mark Bun and Bo Waggoner

2012-12-03

Outline

Recap

Peer prediction:

- Elicit?
- Rewards?

Recap

Peer prediction:

- Elicit?
- Rewards?

Bayesian Truth Serum:

- Elicit?
- Rewards?

Big problem with implementing PP?

Big problem with implementing PP?

- Mechanism needs to know the information structure!

Big problem with implementing PP?

- Mechanism needs to know the information structure!

Big problem with BTS?

Big problem with implementing PP?

- Mechanism needs to know the information structure!

Big problem with BTS?

- Requires $n \to \infty$!

Big problem with implementing PP?

- Mechanism needs to know the information structure!

Big problem with BTS?

- Requires $n \to \infty$!

Other problems with BTS? Does RBTS resolve them?

Goal: Truthful mechanism for any n that doesn't rely on the mechanism knowing the information structure.

Goal: Truthful mechanism for any n that doesn't rely on the mechanism knowing the information structure.

Approach:

- Recall: Elicit two-part report (prediction and signal)
- Which one is easy to incentivize and which is difficult?

Goal: Truthful mechanism for any n that doesn't rely on the mechanism knowing the information structure.

Approach:

- Recall: Elicit two-part report (prediction and signal)
- Which one is easy to incentivize and which is difficult?

To elicit the signal: Use shadowing!

Belief system

Belief system

- Consists of state T drawn from $\{1, \ldots, m\}$ and signals S drawn from $\{l, h\} = \{0, 1\}$

Belief system

- Consists of state T drawn from $\{1, \ldots, m\}$ and signals S drawn from $\{l, h\} = \{0, 1\}$
- Agents have common prior $\Pr[T = t]$ and $\Pr[S = h | T = t]$.

Belief system

- Consists of state T drawn from $\{1, \ldots, m\}$ and signals S drawn from $\{l, h\} = \{0, 1\}$
- Agents have common prior $\Pr[T = t]$ and $\Pr[S = h | T = t]$.
- "Impersonally informative"

Belief system

- Consists of state T drawn from $\{1, \ldots, m\}$ and signals S drawn from $\{l, h\} = \{0, 1\}$
- Agents have common prior $\Pr[T = t]$ and $\Pr[S = h | T = t]$.
- "Impersonally informative" : For every i, j, k, write

$$p_{\{s_i\}} = \Pr[S_j = h | S_i = s_i]$$

Belief system

- Consists of state T drawn from $\{1, \ldots, m\}$ and signals S drawn from $\{l, h\} = \{0, 1\}$
- Agents have common prior $\Pr[T = t]$ and $\Pr[S = h | T = t]$.
- "Impersonally informative" : For every i, j, k, write

$$p_{\{s_i\}} = \Pr[S_j = h | S_i = s_i]$$
$$p_{\{s_i, s_j\}} = \Pr[S_k = h | S_i = s_i, S_j = s_j]$$

Admissibility

Admissibility

Admissible prior:

Admissibility

Admissible prior:

1. At least two states ($m \geq 2$)

Admissibility

Admissible prior:

1. At least two states ($m \geq 2$)
2. Every state has positive probability ($\Pr[T = t] > 0$)

Admissibility

Admissible prior:

1. At least two states ($m \geq 2$)
2. Every state has positive probability ($\Pr[T = t] > 0$)
3. Assortative property:
   $\Pr[S = h | T = 1] < \cdots < \Pr[S = h | T = m]$

Admissibility

Admissible prior:

1. At least two states ($m \geq 2$)
2. Every state has positive probability ($\Pr[T = t] > 0$)
3. Assortative property:
   $\Pr[S = h | T = 1] < \cdots < \Pr[S = h | T = m]$
4. Fully mixed: $0 < \Pr[S = h | T = t] < 1$

Admissibility

Admissible prior:

1. At least two states ($m \geq 2$)
2. Every state has positive probability ($\Pr[T = t] > 0$)
3. Assortative property:
   $\Pr[S = h | T = 1] < \cdots < \Pr[S = h | T = m]$
4. Fully mixed: $0 < \Pr[S = h | T = t] < 1$

Lemma 6: For an admissible prior,

$$1 > p_{\{h,h\}} > p_{\{h\}} > p_{\{h,l\}} = p_{\{l,h\}} > p_{\{l\}} > p_{\{l,l\}} > 0$$

What does shadowing solve?

What does shadowing solve?

- Truthfully elicit *signals* instead of *beliefs*

What does shadowing solve?

- Truthfully elicit *signals* instead of *beliefs*
- $\omega \in \{0, 1\}$ a binary future event

What does shadowing solve?

- Truthfully elicit *signals* instead of *beliefs*
- $\omega \in \{0, 1\}$ a binary future event
- Agent i draws a signal $S_i \in \{0, 1\} = \{l, h\}$

What does shadowing solve?

- Truthfully elicit *signals* instead of *beliefs*
- $\omega \in \{0, 1\}$ a binary future event
- Agent i draws a signal $S_i \in \{0, 1\} = \{l, h\}$
- How do we get agent i to truthfully reveal S_i?

Naïve approach

Naïve approach

1. Choose a strictly proper binary scoring rule $R(y, \omega)$

Naïve approach

1. Choose a strictly proper binary scoring rule $R(y, \omega)$
2. Ask agent i for signal report $x_i \in \{0, 1\}$ using $R(x_i, \omega)$

Naïve approach

1. Choose a strictly proper binary scoring rule $R(y, \omega)$
2. Ask agent i for signal report $x_i \in \{0, 1\}$ using $R(x_i, \omega)$

Might not work if i's beliefs $p_{\{S_i\}}$ about ω are far from x_i!

Solution: Use a reference prediction

Solution: Use a reference prediction

1. Let $y \in (0, 1)$ be an arbitrary prediction report

Solution: Use a reference prediction

1. Let $y \in (0, 1)$ be an arbitrary prediction report
2. Set $\delta = \min(y, 1 - y)$

Solution: Use a reference prediction

1. Let $y \in (0, 1)$ be an arbitrary prediction report
2. Set $\delta = \min(y, 1 - y)$
3. Create a shadow report for i:

$$y_i' = \begin{cases} y + \delta & \text{if} \quad x_i = 1 \\ y - \delta & \text{if} \quad x_i = 0 \end{cases}$$

Solution: Use a reference prediction

1. Let $y \in (0,1)$ be an arbitrary prediction report
2. Set $\delta = \min(y, 1-y)$
3. Create a shadow report for i:

$$y_i' = \begin{cases} y + \delta & \text{if} \quad x_i = 1 \\ y - \delta & \text{if} \quad x_i = 0 \end{cases}$$

4. Score agent x_i using the quadratic scoring rule $R_q(y_i', \omega)$

When and why does shadowing work?

When and why does shadowing work?

- Intuition: Truthfully reporting $x_i = S_i$ pulls the reference prediction toward i's posterior

When and why does shadowing work?

- Intuition: Truthfully reporting $x_i = S_i$ pulls the reference prediction toward i's posterior
- Selten '98: Let $Y \subset [0, 1]$. If agent i has beliefs p, then she maximizes her expected score $R_q(y, \omega)$ over $y \in Y$ by minimizing $|y - p|$.

When and why does shadowing work?

- Intuition: Truthfully reporting $x_i = S_i$ pulls the reference prediction toward i's posterior
- Selten '98: Let $Y \subset [0, 1]$. If agent i has beliefs p, then she maximizes her expected score $R_q(y, \omega)$ over $y \in Y$ by minimizing $|y - p|$.
- **Lemma 8**: Suppose agent i has observed signals $\mathcal{I} \in \{l, h\}^k$ and S_i. If $p_{\{l,\mathcal{I}\}} < y < p_{\{h,\mathcal{I}\}}$, then agent i should truthfully report $x_i = S_i$.

What's missing?

What's missing?

- How do we pick y to guarantee $p_{\{l,\mathcal{I}\}} < y < p_{\{h,\mathcal{I}\}}$?

What's missing?

- How do we pick y to guarantee $p_{\{l,\mathcal{I}\}} < y < p_{\{h,\mathcal{I}\}}$?
- How do we pick ω when there is no ground truth?

Robust Bayesian Truth Serum

Robust Bayesian Truth Serum

- Idea: Shadowing + reference raters

Robust Bayesian Truth Serum

- Idea: Shadowing + reference raters
- **Information report:** $x_i \in \{0, 1\}$ represents agent i's signal

Robust Bayesian Truth Serum

- Idea: Shadowing + reference raters
- **Information report:** $x_i \in \{0, 1\}$ represents agent i's signal
- **Prediction report:** $y_i \in [0, 1]$ represents agent i's prediction for the frequency of h signals

Scoring

Scoring

- For each agent i, pick reference raters j, k

Scoring

- For each agent i, pick reference raters j, k
- Set $y = y_j$ and $\omega = x_k$.

Scoring

- For each agent i, pick reference raters j, k
- Set $y = y_j$ and $\omega = x_k$.
- Create a shadow report for i:

$$y_i' = \begin{cases} y_j + \delta & \text{if} \quad x_i = 1 \\ y_j - \delta & \text{if} \quad x_i = 0 \end{cases}$$

where $\delta = \min(y_j, 1 - y_j)$

Scoring

- For each agent i, pick reference raters j, k
- Set $y = y_j$ and $\omega = x_k$.
- Create a shadow report for i:

$$y_i' = \begin{cases} y_j + \delta & \text{if} \quad x_i = 1 \\ y_j - \delta & \text{if} \quad x_i = 0 \end{cases}$$

where $\delta = \min(y_j, 1 - y_j)$

- **RBTS score:**

$$\underbrace{R_q(y_i', x_k)}_{\text{information score}} \quad + \quad \underbrace{R_q(y_i, x_k)}_{\text{prediction score}}$$

Incentive compatibility

Incentive compatibility

- If $n \geq 3$ agents have signals drawn from an **admissible** common prior, then reporting truthfully is a **strict Bayes-Nash equilibrium**.

Incentive compatibility

- If $n \geq 3$ agents have signals drawn from an **admissible** common prior, then reporting truthfully is a **strict Bayes-Nash equilibrium**.
- Why should agents report their predictions truthfully?

Incentive compatibility

- If $n \geq 3$ agents have signals drawn from an **admissible** common prior, then reporting truthfully is a **strict Bayes-Nash equilibrium**.
- Why should agents report their predictions truthfully?
- How about signals?

Incentive compatibility

- If $n \geq 3$ agents have signals drawn from an **admissible** common prior, then reporting truthfully is a **strict Bayes-Nash equilibrium**.
- Why should agents report their predictions truthfully?
- How about signals? Enough to show that admissibility $\implies p_{\{l,\mathcal{I}\}} < y_j < p_{\{h,\mathcal{I}\}}$.

Incentive compatibility

Incentive compatibility

- **Lemma 6:** For an admissible prior,

$$1 > p_{\{h,h\}} > p_{\{h\}} > p_{\{h,l\}} = p_{\{l,h\}} > p_{\{l\}} > p_{\{l,l\}} > 0$$

Incentive compatibility

- **Lemma 6:** For an admissible prior,

$$1 > p_{\{h,h\}} > p_{\{h\}} > p_{\{h,l\}} = p_{\{l,h\}} > p_{\{l\}} > p_{\{l,l\}} > 0$$

- **Case 1:** $y_j = p_{\{h\}} \implies p_{\{l,h\}} < y_j = p_{\{h\}} < p_{\{h,h\}}$

Incentive compatibility

- **Lemma 6:** For an admissible prior,

$$1 > p_{\{h,h\}} > p_{\{h\}} > p_{\{h,l\}} = p_{\{l,h\}} > p_{\{l\}} > p_{\{l,l\}} > 0$$

- **Case 1:** $y_j = p_{\{h\}} \implies p_{\{l,h\}} < y_j = p_{\{h\}} < p_{\{h,h\}}$
- **Case 2:** $y_j = p_{\{l\}} \implies p_{\{l,l\}} < y_j = p_{\{l\}} < p_{\{h,l\}}$

Incentive compatibility

- **Lemma 6:** For an admissible prior,

$$1 > p_{\{h,h\}} > p_{\{h\}} > p_{\{h,l\}} = p_{\{l,h\}} > p_{\{l\}} > p_{\{l,l\}} > 0$$

- **Case 1:** $y_j = p_{\{h\}} \implies p_{\{l,h\}} < y_j = p_{\{h\}} < p_{\{h,h\}}$
- **Case 2:** $y_j = p_{\{l\}} \implies p_{\{l,l\}} < y_j = p_{\{l\}} < p_{\{h,l\}}$
- Recap: $p_{\{l,s_j\}} < y_j < p_{\{h,s_j\}}$, so agent i should truthfully reveal $x_i = S_i$

Properties of RBTS

Properties of RBTS

- Works for any number of agents $n \geq 3$

Properties of RBTS

- Works for any number of agents $n \geq 3$
- Scores are well-defined between 0 and 2.

Properties of RBTS

- Works for any number of agents $n \geq 3$
- Scores are well-defined between 0 and 2. Participation is "ex post individually rational".

Properties of RBTS

- Works for any number of agents $n \geq 3$
- Scores are well-defined between 0 and 2. Participation is "ex post individually rational".
- Numerically robust

Extensions

Extensions

- Adapting to a budget constraint.

Extensions

- Adapting to a budget constraint. Scaling, randomized exclusion.

Extensions

- Adapting to a budget constraint. Scaling, randomized exclusion.
- More than two outcomes?

Extensions

- Adapting to a budget constraint. Scaling, randomized exclusion.
- More than two outcomes?
- Other proper scoring rules?

Peer Prediction Without a Common Prior. Witkowski, Parkes. EC-2012.

Peer Prediction Without a Common Prior. Witkowski, Parkes. EC-2012.

- Same setting as RBTS except each agent may have a different information structure.

Peer Prediction Without a Common Prior. Witkowski, Parkes. EC-2012.

- Same setting as RBTS except each agent may have a different information structure.
- Elicit both the agent's prior *and* posterior probability that another agent will get a high signal.

Peer Prediction Without a Common Prior. Witkowski, Parkes. EC-2012.

- Same setting as RBTS except each agent may have a different information structure.
- Elicit both the agent's prior *and* posterior probability that another agent will get a high signal.
- Infer whether agent's signal was high or low

Peer Prediction Without a Common Prior. Witkowski, Parkes. EC-2012.

- Same setting as RBTS except each agent may have a different information structure.
- Elicit both the agent's prior *and* posterior probability that another agent will get a high signal.
- Infer whether agent's signal was high or low
- Score = score(prior) + score(posterior)

Peer Prediction Without a Common Prior. Witkowski, Parkes. EC-2012.

- Same setting as RBTS except each agent may have a different information structure.
- Elicit both the agent's prior *and* posterior probability that another agent will get a high signal.
- Infer whether agent's signal was high or low
- Score = score(prior) + score(posterior)
- OR elicit prior and signal, and shadow to get a posterior (complicated)

	PP	PP-CP	BTS	RBTS
common prior exists				
designer knows CP				
information structure				
# of players				
# of signals				
designer learns?				