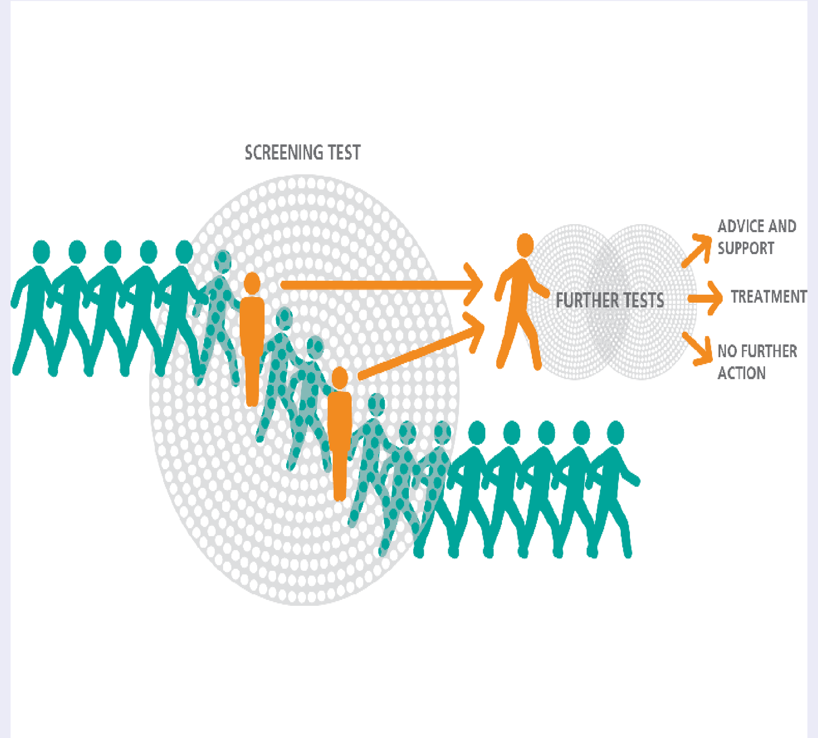
Optimal Questionnaires for Screening of Strategic Agents

Anuj S. Vora, Ankur A. Kulkarni (Systems and Control Engineering, IIT Bombay)
anujvora@iitb.ac.in, kulkarni.ankur@iitb.ac.in

Paper: 5278

Background

- Health inspector encounters travellers



- Limited resources, not all travellers tested
- Identify susceptible travellers from their travel history
- However, people have a tendency to misreport their true travel history, due to stigma, inconvenience due to testing and quarantine protocols

Model

- $\mathcal{X}$ - locations, a location is COVID-19 hotspot or safe, $|\mathcal{X}| < \infty$
- $x \in \mathcal{X}^n$ - travel history, a sequence of n locations.
- λ - type of traveller, degree of honesty, $|\Lambda| < \infty$.
- $\mathbb{P}_\Lambda$ - probability distribution of types, belief of inspector
- We term the health inspector as the receiver and the travellers as senders

Strategies of the senders and the receiver

- The receiver presents travellers with a common questionnaire
- $\mathcal{C}^n \subseteq \mathcal{X}^n$ - check-list, and travellers have to choose exactly one
- Sender λ answers as $s_n^\lambda : \mathcal{X}^n \to \mathcal{C}^n$
- Receiver interprets as $g_n : \mathcal{C}^n \to \mathcal{X}^n$
- Sufficient to choose a strategy g_n where $g_n(x) = x \ \forall\ x \in \mathcal{C}^n$.

Objectives of the inspector and travellers

- Set of perfectly recovered sequences
$$\mathcal{D}(g_n, s_n^\lambda) := \left\{x \in \mathcal{X}^n \mid g_n \circ s_n^\lambda(x) = x\right\}$$
- x - true history, $\widehat{x}$ - recovered history, then sender λ gets
$$\mathscr{U}_n(\widehat{x}, x, \lambda) = \frac{1}{n}\sum_{i=1}^{n} \mathscr{U}(\widehat{x}_i, x_i, \lambda) \qquad \forall\ x, \widehat{x} \in \mathcal{X}^n,$$
with $\mathscr{U} : \mathcal{X} \times \mathcal{X} \times \Lambda \to \mathbb{R}$ being the single letter utility.
- Receiver maximizes $\sum_{\lambda\in\Lambda} \mathbb{P}(\lambda)|\mathcal{D}(g_n, s_n^\lambda)|$
- Senders maximizes $\mathscr{U}(g_n \circ s_n^\lambda(x), x, \lambda)$

Definitions

- **Sender graph** - Graph $G_\lambda^n = (\mathcal{X}^n, E)$, where $(x, y) \in E$ if $\mathscr{U}_n(x, x, \lambda) \le \mathscr{U}_n(y, x, \lambda)$ or $\mathscr{U}_n(y, y, \lambda) \le \mathscr{U}_n(x, y, \lambda)$.
- **λ-partition of a set** - $\bar{I}_\lambda^n$ is the largest subset of I^n which is an independent set in G_λ^n
- **Rate of information extraction** - For a strategy g_n,
$$R(g_n) = \left(\sum_{\lambda\in\Lambda} \mathbb{P}(\lambda)\left(\min_{s_n^\lambda \in \mathscr{B}(g_n,\lambda)} |\mathcal{D}(g_n, s_n^\lambda)|\right)\right)^{1/n}.$$

Formulation of problem as a Stackelberg game

- We consider a Stackelberg game with the receiver as the leader
- In a Stackelberg equilibrium, the strategies are
$$g_n^* \in \arg\max_{g_n} \sum_{\lambda\in\Lambda} \mathbb{P}(\lambda)\left(\min_{s_n^\lambda \in \mathscr{B}(g_n,\lambda)} |\mathcal{D}(g_n, s_n^\lambda)|\right),$$
where the best response set of the sender λ is $\mathscr{B}(g_n, \lambda)$, where
$$\mathscr{B}(g_n, \lambda) = \Big\{s_n^\lambda : \mathcal{X}^n \to \mathcal{X}^n \mid \mathscr{U}_n(g_n \circ s_n^\lambda(x), x, \lambda) \ge \mathscr{U}_n(g_n \circ s_n'^\lambda(x), x, \lambda) \forall\ x \in \mathcal{X}^n, \forall\ s_n'^\lambda\Big\}.$$
- Assuming a pessimistic receiver, we incorporate the minimization over $\mathscr{B}(g_n, \lambda)$.

Example

- Let $\mathcal{X} = \{0, 1, 2\}$ and $\Lambda = \{h, d\}$, where h: *honest*, d: *dishonest*. Let $\mathbb{P}_\Lambda(h) = 1/3$ and $\mathbb{P}_\Lambda(d) = 2/3$.
- Let the utility of the sender h be $\mathscr{U}(x, x, h) > \mathscr{U}(x', x, h) \ \ \forall\ x' \in \mathcal{X}, x' \ne x$.
- For the sender type $\lambda = d$, the utility is
$$\mathscr{U}(0,0,d) = 0,\ \mathscr{U}(1,0,d) = 1,\ \mathscr{U}(2,0,d) = -1,$$
$$\mathscr{U}(0,1,d) = 1,\ \mathscr{U}(1,1,d) = 0,\ \mathscr{U}(2,1,d) = -1,$$
$$\mathscr{U}(0,2,d) = 1,\ \mathscr{U}(1,2,d) = 1,\ \ \mathscr{U}(2,2,d) = 0.$$
- Let $n = 1$. Suppose $g(i) = i \ \forall\ i \in \mathcal{X}$, i.e., $\mathcal{C} = \{0, 1, 2\}$
- The best response of the sender h is $s^h(x) = x$ for all $x \in \mathcal{X}$. For the sender type d, $\mathscr{B}(g, d) = \{\bar{s}^d, \bar{\bar{s}}^d\}$, where
$$\bar{s}^d(i) = \begin{cases} 1 & i = 0 \\ 0 & i = 1 \\ 0 & i = 2 \end{cases}, \quad \bar{\bar{s}}^d(i) = \begin{cases} 1 & i = 0 \\ 0 & i = 1 \\ 1 & i = 2 \end{cases}.$$
- The set of perfectly recovered sequences are $\mathcal{D}(g, s^h) = \mathcal{X}$ and $\mathcal{D}(g, s^d) = \varnothing$ for all $s^d \in \mathscr{B}(g, d)$ respectively. This gives
$$\sum_{\lambda\in\Lambda} \mathbb{P}(\lambda)\left(\min_{s^\lambda \in \mathscr{B}(g,\lambda)} |\mathcal{D}(g, s^\lambda)|\right) = \frac{1}{3}3 + \frac{2}{3}0 = 1.$$

- Suppose instead the receiver chooses a strategy $\widetilde{g}$ as
$$\widetilde{g}(i) = \begin{cases} 0 & i = 0 \\ 0 & i = 1 \\ 2 & i = 2 \end{cases}.$$
Thus, $\mathcal{C} = \{0, 2\}$
- Again, for the sender h, the best response strategy is $\widetilde{s}^h(x) = x \ \forall\ x \in \mathcal{X}$. For the sender d, the best response strategy is $\widetilde{s}^d(i) = 0 \ \forall\ i \in \mathcal{X}$.
- Now $\mathcal{D}(g, \widetilde{s}^h) = \{0, 2\}$ and $\mathcal{D}(g, \widetilde{s}^d) = \{0\}$ and hence
$$\sum_{\lambda\in\Lambda} \mathbb{P}(\lambda)\left(\min_{s^\lambda \in \mathscr{B}(\widetilde{g},\lambda)} |\mathcal{D}(\widetilde{g}, s^\lambda)|\right) = \frac{1}{3}2 + \frac{2}{3}1 = \frac{4}{3}.$$

Results

- For all equilibrium strategies g_n^* of the receiver
$$\sum_{\lambda\in\Lambda} \mathbb{P}(\lambda)\left(\min_{s_n^\lambda \in \mathscr{B}(g_n^*,\lambda)} |\mathcal{D}(g_n^*, s_n^\lambda)|\right) = \max_{I^n \subseteq \mathcal{X}^n} \sum_{\lambda\in\Lambda} \mathbb{P}(\lambda)|\bar{I}_\lambda^n|,$$
where $\{\bar{I}_\lambda^n\}_{\lambda\in\Lambda}$ is the collection of the λ-partitions of I^n.
- For all Stackelberg equilibrium strategies g_n^*,
$$\alpha\left(\cup_\lambda G_\lambda^n\right)^{1/n} \le R(g_n^*) \le \left(\sum_{\lambda\in\Lambda} \mathbb{P}_\Lambda(\lambda)\alpha(G_\lambda^n)\right)^{1/n}.$$
- For all sequences of Stackelberg equilibrium strategies $\{g_n^*\}_{n\ge1}$ of the receiver,
$$\alpha\left(\cup_\lambda G_\lambda\right) \le \limsup_{n\to\infty} R(g_n^*) \le \Xi(\mathscr{U}, \lambda^*),$$
where $\lambda^* = \max_{\lambda\in\Lambda} \Xi(\mathscr{U}, \lambda)$, $\Xi(\mathscr{U}, \lambda) := \lim_{n\to\infty} \alpha(G_\lambda^n)^{1/n}$.
- The bounds depend only on the single letter utility $\mathscr{U}(\cdot,\cdot,\lambda)$, fundamental bounds on the rate

Main insight

To get correct information, ask neither too many questions nor too less questions

Conclusion

- We posed and studied a problem of information extraction from strategic
- Interesting observations even for a simple type of questionnaire
- We characterized the optimal questionnaires for the receiver
- We also derived bounds on the rate of information extraction, for finite n and for n growing to infinity