CSCI699: Topics in Learning and Game Theory

Lecture 11

Lecturer: Shaddin Dughmi *Scribes: Brendan Avent and Alana Shine*

Recall:

- State of nature $\omega \in \Omega$ drawn from prior $P(\omega)$.
- n agents where agent i receives a signal $t_i \in T$ distributed according to $p(t|\omega)$.
- We assume that t_i and t_j are independent given ω.
- Our goal is to incentivize truthful reporting of $t_1, t_2, ..., t_n$.

We analyze incentive design under three different frameworks.

1. Assumption 1: Type reports t_i today, after get access to state ω, at which point I decide payments. This reduces to assigning rewards with a proper scoring rule.
2. Assumption 2: Never get state of nature, but I get $p(\omega)$. We call this peer prediction.
3. Assumption 3: Bayesian Truth-Serum. No knowledge of ω or $p(\omega)$, just the assumption that agents act rationally.

Bayesian Truth Serum Protocol

- Let $S : \Delta(T) \times T$ be a strictly proper scoring rule.
- Solicit (1) a report $r_i \in T$ and (2) a prediction $y_i \in \Delta(T)$.
- For each type $t \in T$, let $\overline{x}_t$ be a fraction of agents reporting t and we write $\overline{x} \in \Delta(T)$ be the probability distribution over types reporting $t \in T$.
- Let $\overline{y} \in [0, 1]^T$ be a geometric mean of predictions.

$$\log \overline{y}_t = \frac{1}{n} \sum_{i=1}^{n} \log y_t^i$$

- Each agent i reporting r_i gets paid

$$\log \frac{\overline{x}_{r_i}}{\overline{y}_{r_i}} + \mathbb{E}_{t\sim\overline{x}}[s(y_i, t)]$$

The first part of the sum, $\log \frac{\overline{x}_{r_i}}{\overline{y}_{r_i}}$ is referred to as the *info score*. The second part of the sum $\mathbb{E}_{t\sim\overline{x}}[s(y_i, t)]$ is derived from choice of proper scoring rule s. We note that the payment does not depend on p or ω.

We need to show that the agent's best response is to provide the true report of it's type t_i and prediction y_i to prove the following theorem.

Theorem 1. *In the BTS protocol, it is a BNE for each agent to report $r_i = t_i$ and $y^i_t = P[t \mid t_i]$.*

Note the following notation:

- $P[t \mid \omega] :=$ the probability of type t in state-of-nature ω
- $P[t \mid t'] :=$ the probability of agent 2 having type t given agent 1 has t'

So by Bayes' Rule, we have:

$$P[\omega \mid t] = \frac{P[\omega] \cdot P[t \mid \omega]}{\sum_{\omega'} P[\omega'] \cdot P[t \mid \omega']}$$

and

$$P[\omega \mid t, t'] = \frac{P[\omega] \cdot P[t \mid \omega] \cdot P[t' \mid \omega]}{\sum_{\omega'} P[\omega'] \cdot P[t \mid \omega'] \cdot P[t' \mid \omega']}.$$

When we say that "agent i reports truthfully", this means that $r_i = t_i$ and $y^i_t = P[t \mid t_i]$.

One easy fact is that if all (or all but 1) agents report truthfully, then for each fixed state-of-nature ω,

$$\begin{aligned} \bar{x}_t &= \frac{1}{n}|\{i : t_i = t\} \\ &\underset{n\to\infty}{=} P[t \mid \omega], \end{aligned}$$

and

$$\begin{aligned}\log \bar{y}_t &= \frac{1}{n}\sum_i y_i \\ &= \sum_{\hat{t}\in T} \bar{x}_{\hat{t}} \log P[t \mid \hat{t}] \\ &\underset{n\to\infty}{=} \sum_{\hat{t}\in T} P[\hat{t} \mid \omega] \log P[t \mid \hat{t}].\end{aligned}$$

Lemma 2. *Suppose all $i' \neq i$ report truthfully. Then it's strictly optimal for i to report $y^i_t = P[t \mid t_i]$ for all t.*

Proof:

- Choice of y^i can't influence the information score.
- The second term, $\underset{t\sim\bar{x}}{\mathbb{E}}[s(y^i, t)]$ is influencable though.
- Given t_i, i believes each other t has $P[t \mid t_i]$.
- Given infinitely many agents, i believes $\bar{x}_t = P[t \mid t_i]$.
- By SPSR (strictly proper scoring rule), $y^i_t = \bar{x}_t = P[t \mid t_i]$ for all t is strictly optimal. □

Another easy fact is if $A \perp B \mid C$, then

$$\frac{P[A \mid C]}{P[A \mid B]} = \frac{P[C \mid A, B]}{P[C \mid B]}.$$

Lemma 3. *Suppose all agents report truthfully. Then it is optimal for agent i to report $r_i = t_i$.*

Proof:

- r_i only affects the information score.
- Reporting $r_i = t$ yields $\log \frac{\bar{x}_t}{\bar{y}_t}$.
- Suppose $t_i = t'$, the expected utility of i for reporting t is

$$\begin{aligned}
\mathbb{E}\left[\log\frac{\bar{x}_t}{\bar{y}_t}\right] &= \sum_{\omega} P[\omega \mid t'] \cdot \mathbb{E}\left[\log\frac{\bar{x}_t}{\bar{y}_t} \mid \omega\right] \\
&\underset{\text{ez fact}}{=} \sum_{\omega} P[\omega \mid t'] \cdot \left(\log P[t \mid \omega] - \sum_{\hat{t}} P[\hat{t} \mid \omega] \cdot \log P[t \mid \hat{t}]\right) \\
&= \sum_{\omega} P[\omega \mid t'] \cdot \sum_{\hat{t}} P[\hat{t} \mid \omega] \cdot \log\frac{P[t \mid \omega]}{P[t \mid t']} \\
&= \sum_{\omega,\hat{t}|t'} P[\omega, \hat{t} \mid t'] \cdot \log\frac{P[t \mid \omega]}{P[t \mid \hat{t}]} \\
&= \sum_{\hat{t}} P[\hat{t} \mid t'] \cdot \sum_{\omega} P[\omega \mid \hat{t}, t'] \cdot \log\frac{P[t \mid \omega]}{P[t \mid \hat{t}]} \\
&= \mathop{\mathbb{E}}_{\hat{t}|t'} \mathop{\mathbb{E}}_{\omega|\hat{t},t'} \log\frac{P[t \mid \omega]}{P[t \mid \hat{t}]} \\
&= \mathop{\mathbb{E}}_{\hat{t}|t'} \mathop{\mathbb{E}}_{\omega|\hat{t},t'} \log\frac{P[\omega \mid \hat{t}, t]}{P[\omega \mid \hat{t}]}.
\end{aligned}$$

So the expected info score if agent i reports some type t is

$$\mathop{\mathbb{E}}_{\hat{t}|t'} \mathop{\mathbb{E}}_{\omega|\hat{t},t'} \log\frac{P[\omega \mid \hat{t}, t]}{P[\omega \mid \hat{t}]}$$

And by using the same analysis, if agent i reports it's true type t' the expected info score is

$$\mathop{\mathbb{E}}_{\hat{t}|t'} \mathop{\mathbb{E}}_{\omega|\hat{t},t'} \log\frac{P[\omega \mid \hat{t}, t']}{P[\omega \mid \hat{t}]}$$

Comparing the expected info score when reporting t to that when reporting t' by subtracting the expected info score given agent i reports t and the expected info score given agent i reports t' to show that reporting t' (the truth) is indeed an optimal strategy for agent i.

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\hat{t}|t'} \mathop{\mathbb{E}}_{\omega|\hat{t},t'} \log \frac{P[\omega \mid \hat{t}, t]}{P[\omega \mid \hat{t}, t']} \underset{\text{Jensen's}}{\leq} & \mathop{\mathbb{E}}_{\hat{t}|t'} \log \mathop{\mathbb{E}}_{\omega|\hat{t},t'} \frac{P[\omega \mid \hat{t}, t]}{P[\omega \mid \hat{t}, t']} \\
= & \mathop{\mathbb{E}}_{\hat{t}|t'} \log \sum_{\omega} P[\omega \mid \hat{t}, t'] \cdot \frac{P[\omega \mid \hat{t}, t]}{P[\omega \mid \hat{t}, t']} \\
= & \mathop{\mathbb{E}}_{\hat{t}|t'} \log 1 \\
= & \, 0. \quad \square
\end{aligned}
$$

Thus, Theorem 1 is directly proven by Lemmas 2 and 3.