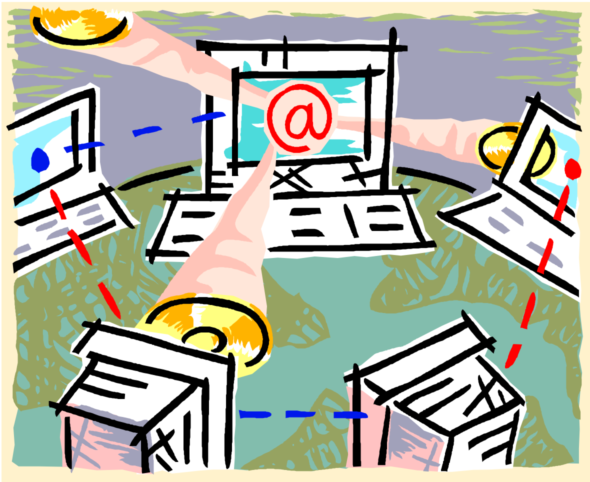
Probability Models of Information Exchange on Networks

Lecture 3



Elchanan Mossel
UC Berkeley

The Bayesian View of the Jury Theorem

- Recall: we assume 0/1 with prior probability (0.5,0.5).

- Each voter receives signal x_i which is correct with probability p independently.

- Note that if this is indeed the case, then after the vote has been cast, all voters can calculate:
- $P[s = 1 \mid x] / P[s = 0 \mid x]$.

- Obtain posterior probability of 1,0.

- Everybody agree about the posterior.

- Can this be extended to networks?

Bayesian Exchange on Networks

- <u>Setup:</u>
- $S \in \{0,1\}$ with $P[S = 1] = ½$ (apriori).
- Distributions of signals D_0, D_1
- A (directed) social network $G = (V,E)$ of size n with self loops.

Bayesian Exchange on Networks

- Setup:
- $S \in \{0,1\}$ with $P[S = 1] = ½$ (apriori).
- Distributions of signals D_0, D_1
- A (directed) social network $G = (V,E)$ of size n with self loops.
- At time 0: agents receive ind. signals $X(i,0)$ from D_S
- Let $F(i,0) = \sigma(X(i))$
- At each discrete time step $t \geq 1$:
- Agent i declares
- $X(i,t) = P[S = 1 | F(i,t-1)] = E[S | F(i,t-1)]$
- Let $F(i,t) = \sigma(X(j,s) : (i \rightarrow j) \in E, s \leq t)$

Bayesian Exchange on Networks

- Setup:
- $S \in \{0,1\}$ with $P[S = 1] = ½$ (apriori).
- Distributions of signals D_0, D_1
- A (directed) social network $G = (V,E)$ of size n with self loops.
- At time 0: agents receive ind. signals $X(i,0)$ from D_S
- Let $F(i,0) = \sigma(X(i))$
- At each discrete time step $t \geq 1$:
- Agent i declares $X(i,t) = E[S \mid F(i,t-1)]$
- Let $F(i,t) = \sigma(X(j,s) : (i \to j) \in E, s \leq t)$

- Q1: Do agents converge?
- Q2: If they do, what to do they converge to?

Convergence of a Single Agent

- At time 0: agent i, receives a signal X(i,0) from D_S
- Let F(i,0) = σ(X(i))
- At each discrete time step t $\geq$ 1:
- Agent i declares X(i,t) = E[S | F(i,t-1)]
- Let F(i,t) = σ(X(j,s) : (i $\rightarrow$ j) $\in$ E, s $\leq$ t)

- Claim: Each agent converges.

Convergence of a Single Agent

- At time 0: agent i, receives a signal X(i,0) from D_S
- Let F(i,0) = σ(X(i))
- At each discrete time step t $\geq$ 1:
- Agent i declares X(i,t) = E[S | F(i,t-1)]
- Let F(i,t) = σ(X(j,s) : (i $\rightarrow$ j) $\in$ E, s $\leq$ t)

- Claim: Each agent converges.
- Pf: X(i,t) is a bounded martingale.

Convergence of a Single Agent

- At time 0: agent i, receives a signal X(i,0) from D_S
- Let F(i.0) = σ(X(i))
- At each discrete time step t $\geq$ 1:
- Agent i declares X(i,t) = E[S | F(i,t-1)]
- Let F(i,t) = σ(X(j,s) : (i $\rightarrow$ j) $\in$ E, s $\leq$ t)

- Claim: Each agent converges.
- Pf: X(i,t) is a bounded martingale.

- Note: Doesn't use anything about
- Network structure or
- Independence of signals.

- Q: Do agents agree in the limit?

Bayesian Exchange on Networks

- Q1: Do agents converge to the same belief?
- Not necessarily.
- For example – disconnected graph.

Bayesian Exchange on Networks

- Q1: Do agents converge to the same belief?
- Not necessarily.
- For example – disconnected graph.
- Or even graph that is not strongly connected.

Agreement

- Q1: Do agents converge to the same belief?
- Not necessarily.
- For example – disconnected graph.
- Or even graph that is not strongly connected.

- Thm (Aumann 76, Geanakoplos & Polemarchakis 82, Parikh, Krasucki):

- If the graph G is strongly connected, all agents will a.s. converge to the same value.

- Recall: Strongly connected means that for every pair of vertices there is a directed path connecting them.

Agreement Proof

Proof Sketch: :

- Let $X(i) = \lim X(i,t) = E[S \mid F(i)]$, where $F(i) = \lim F(i,t)$.

- $X(i)$ is the function closest to S in $L^2(F(i))$.

Agreement Proof

Proof Sketch: :

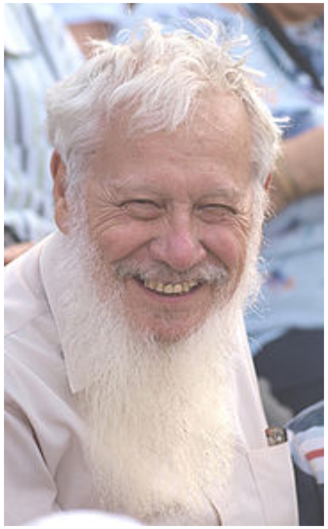
- Let $X(i) = \lim X(i,t) = E[S \mid F(i)]$, where $F(i) = \lim F(i,t)$.
- $X(i)$ is the function closest to S in $L^2(F(i))$.
- If $i \rightarrow j$ in G then $X(j) \in L^2(F(i))$.
- Therefore: $|X(i)-S|_2 \leq |X(j)-S|_2$.
- Strong connectivity $\Rightarrow \forall\, i,j$: $|X(i)-S|_2 = |X(j)-S|_2$

Agreement Proof

Proof Sketch: :

- Let $X(i) = \lim X(i,t) = E[S \mid F(i)]$, where $F(i) = \lim F(i,t)$.
- $X(i)$ is the function closest to S in $L^2(F(v))$.
- If $i \rightarrow j$ in G then $X(j) \in L^2(F(i))$.
- Therefore: $|X(i)-S|_2 \leq |X(j)-S|_2$.
- Strong connectivity $\Rightarrow \forall\, i,j$: $|X(i)-S|_2 = |X(j)-S|_2$
- If $i \rightarrow j$, $P[X(i) \neq X(j)] > 0$ then $Z = 0.5(X(i)+X(j)) \in F(i)$ and Z closer to S than either $X(i)$ or $X(j)$.
- Strong connectivity $\Rightarrow \forall\, i,j$: $X(i) = X(j)$.

Agreement History



• Note: Result and proof above did not use independence of signals.

•Aumann 76: notion of common knowledge:
" If two people have the same priors and their posterior for a given event are common knowledge, then these posteriors must be equal"

• Critique of Bayesian economics.

• Geanakoplos & Polemarchakis 82: Dynamics with two individuals.

• Parikh, Krasucki: Networks

The Learning Problem

- Assume G is strongly connected.
- Do agents learn?
- Strongest possible sense of learning:
- Do as well as if all see all signals.
- Strongest possible sense of non-learning:
- Do not do better than random.
- Consider: dependent / independent signals.

Non-Learning

- Assume G is strongly connected.

- Do agents learn?

- Example of non learning: G = ({1,2}, {(1,2)})
- S(1) and S(2) uniform +/- with S = S(1) S(2)
- X(1,t) = X(2,t) = 0 for all t.

- Or G = K_n where ½ of the vertices get S(1) and ½ get S(2).

- A lot of information but it is all lost.

- How about if the the signals are (conditionally) independent?

Learning with independent signals

- Assume G is strongly connected &
- The signals X(i) are i.i.d. conditionally on S.
- Let X = $\lim_t$ X(i,t) = limit common belief.

- <u>Thm (M, Sly, Tamuz):</u> X = E[S | X(1),…,X(n)]

Learning with independent signals

- Assume G is strongly connected &
- The signals X(i) are i.i.d. conditionally on S.
- Let $X = \lim_t X(i,t)$ = limit common belief.

- Thm (M, Sly, Tamuz): $X = E[S \mid X(1),\ldots,X(n)]$

- Agents aggregate optimally!

- Statement and proof work for any model where the posterior beliefs are common knowledge.

Learning with independent signals

- Assume G is strongly connected &
- The signals X(i) are i.i.d. conditionally on S.
- Let $X = \lim_t X(i,t)$ = limit common belief.

- Thm (M, Sly, Tamuz): $X = E[S \mid X(1),\ldots,X(n)]$

- Agents aggregate optimally!

- Statement and proof work for any model where the posterior beliefs are common knowledge.

- The proof uses Chebyshev's sum inequality: if f and g are strictly increasing then $E[f(X)\, g(X)] \geq E[f(X)]\, E[g(X))]$ and equality means that g = c f.

Agreeing on beliefs implies learning - proof

Belief Learning Theorem (M. Sly and Tamuz (2012))

If there exists a random variable X such that $X = X_i := \mathbb{E}[S \mid \mathcal{F}_i(\infty)]$ for all i then all agents learned optimally:

$$X = \mathbb{P}[S = 1 \mid X(1), \ldots, X(n)] .$$

Proof Sketch

- $$Z_i := \log \frac{\mathbb{P}\left[S = 1 \mid X(i)\right]}{\mathbb{P}\left[S = 0 \mid X(i)i\right]} = \log \frac{\mathbb{P}\left[X(i) \mid S = 1\right]}{\mathbb{P}\left[X(i) \mid S = 0\right]}, \quad Z = \sum_i Z_i$$

  so

  $$\mathbb{P}\left[S = 1 \mid X(1), \ldots, X(n)\right] = L(Z)$$

  where $L(x) = e^x/(e^x + e^{-x})$.

Proof Sketch

- $$Z_i := \log \frac{\mathbb{P}[S = 1 \mid X(i)]}{\mathbb{P}[S = 0 \mid X(i)i]} = \log \frac{\mathbb{P}[X(i) \mid S = 1]}{\mathbb{P}[X(i) \mid S = 0]}, \quad Z = \sum_i Z_i$$

  so

  $$\mathbb{P}[S = 1 \mid X(1), \ldots, X(n)] = L(Z)$$

  where $L(x) = e^x/(e^x + e^{-x})$.
- Since X is $\mathcal{F}_i$ measurable

  $$X = \mathbb{E}[L(Z) \mid \mathcal{F}_i] = \mathbb{E}[L(Z) \mid X].$$

Proof Sketch

- $$Z_i := \log \frac{\mathbb{P}[S=1 \mid X(i)]}{\mathbb{P}[S=0 \mid X(i)i]} = \log \frac{\mathbb{P}[X(i) \mid S=1]}{\mathbb{P}[X(i) \mid S=0]}, \quad Z = \sum_i Z_i$$

  so
  $$\mathbb{P}[S=1 \mid X(1),\dots,X(n)] = L(Z)$$
  where $L(x) = e^x/(e^x + e^{-x})$.
- Since X is $\mathcal{F}_i$ measurable
  $$X = \mathbb{E}[L(Z) \mid \mathcal{F}_i] = \mathbb{E}[L(Z) \mid X].$$
- Hence since Z_i is $\mathcal{F}_i$ measurable
  $$\begin{aligned}\mathbb{E}[Z_i \cdot L(Z) \mid X] &= \mathbb{E}[\mathbb{E}[Z_i \cdot L(Z) \mid \mathcal{F}_i] \mid X] \\ &= \mathbb{E}[Z_i \cdot X \mid X] \\ &= \mathbb{E}[Z_i \mid X]\,\mathbb{E}[L(Z) \mid X]\end{aligned}$$

Proof concluded

$$\mathbb{E}\left[Z_i \cdot L(Z) \mid X\right] = \mathbb{E}\left[Z_i \mid X\right] \mathbb{E}\left[L(Z) \mid X\right]$$

- Summing over i we get that

$$\mathbb{E}\left[Z \cdot L(Z) \mid X\right] = \mathbb{E}\left[Z \mid X\right] \mathbb{E}\left[L(Z) \mid X\right]$$

Proof concluded

$$\mathbb{E}\left[Z_i \cdot L(Z) \mid X\right] = \mathbb{E}\left[Z_i \mid X\right] \mathbb{E}\left[L(Z) \mid X\right]$$

- Summing over i we get that

$$\mathbb{E}\left[Z \cdot L(Z) \mid X\right] = \mathbb{E}\left[Z \mid X\right] \mathbb{E}\left[L(Z) \mid X\right]$$

- Since $L(x)$ is strictly increasing this implies that Z is constant conditional on X, i.e. Z is X measurable so

$$X = \mathbb{E}\left[L(Z) \mid X\right] = \mathbb{E}\left[L(Z) \mid Z\right] = L(Z)$$

- So the agreed value X equal to the optimal estimator $L(Z)$ as needed.

Some Open Problems

- Open problem 1:
- Is the learning theorem valid under weaker conditions on the distributions?

- Open problem 2:
- How quick is the convergence to the agreed value?

- Open problem 3:
- Are there good algorithms to compute the dynamics?

- We will now look at problems 2 and 3 in some simple special cases.

Learning in finite probability spaces

- Question: Assume the probability space of the state of the world and the signals is finite. Does the learning process converge in a finite number of iterations?

Learning in finite probability spaces

- Claim: Consider the process on a graph G with n vertices and assume that the size of the probability space (including S and all signals) is a finite M.
- Then the learning process converges in at most M n iteration.

Learning in finite probability spaces

- Claim: Consider the process on a graph G with n vertices and assume that the size of the probability space Ω (including S and all signals) is a finite M.
- Then the learning process converges in at most M n iteration.

- Pf:
- The information $F_i(t)$ may be encoded by $S_i(t) \subset \Omega$.
- Satisfying: $S_i(t+1) \subseteq S_i(t)$.
- If $S_i(t+1) = S_i(t)$ for some t and all i then the process has converged.
- Argument is close to that of Geanakoplos & Polemarchakis 82.
- Open problem: Can this bound be improved?

An Example of Learning in Finite Spaces

- Example: $X \in [n^2]$ with uniform prior. The signals are:
- Player 1: $X \in \{1…,n\}$, $X \in \{n+1,..,2n\}$ etc.
- Player 2: $X \in \{1,…,n+1\}$,$X \in \{n+2,..,2n+2\}$,…, $X \in \{n^2\}$
- True value is X is sampled to be 1.
- The event the players are estimating
- $S = 1(X \in \{1,n+2,2n+3, 3n + 4,…, n^2 -1, n^2\})$.

- Q: What will happen?

An Example of Learning in Finite Spaces

- Example: $X \in [n^2]$ with uniform prior. The signals are:
- Player 1: $X \in \{1\ldots,n\}$, $X \in \{n+1,..,2n\}$ etc.
- Player 2: $X \in \{1,\ldots,n+1\}$,$X \in \{n+2,..,2n+2\}$,…, $X \in \{n^2\}$
- True value is X is sampled to be 1.
- The event the players are estimating
- $S = 1(X \in \{1,n+2,2n+3, 3n + 4,\ldots, n^2 -1, n^2\})$.

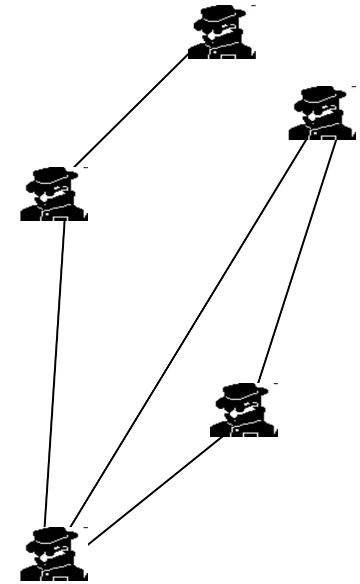
- What will happen?
- Player 1 will say 1/n
- Player 2 will say 1/(n+1)
- Player 1 learns that it is not n^2 but will still say 1/n.
- Player 2 learns that player 1 was not in the last group but will still say 1/(n+1).
- Q: How tight is the bound?

Next Example

- We will talk about a Gaussian model which is:
- Computationally feasible
- Has rapid convergence.
- Converges to the optimal answer for every connected network.

- Following model was studied in P. DeMarzo, D. Vayanos, and J. Zwiebel. and by Mossel and Tamuz.

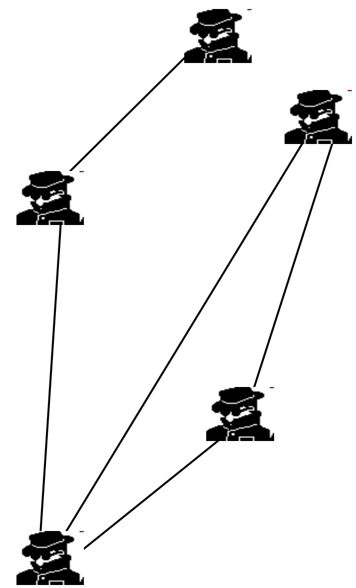
The Gaussian Model



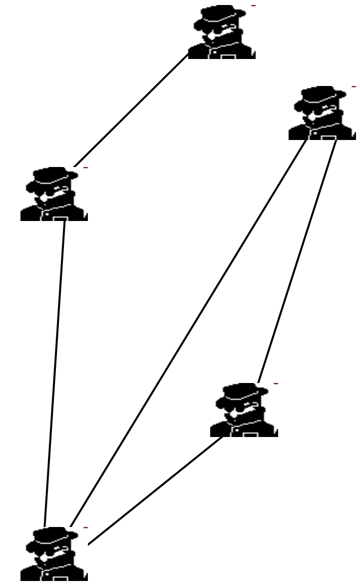
- The original signals are N(μ = ?,1).
- In each iteration
  - Each agent action reveals her current estimate of μ to her neighbors.
  - E.g. set price by min utility $(x- \mu)^2$
  - Each agent calculates a new estimate of μ based on her neighbors' broadcasts.
- Assume agents know the graph structure.
- Repeat *ad infinitum*
- Assume agents know the graph structure.
- Example: interval of length 4.

Utopia

- "Network Learns" **Avg**(X_v)
- Variance of this estimator is 1/n.
- Could be achieved if everyone was friends with everyone.
- Technical comments: This is both the
- ML estimator &
- Bayesian estimator with uniform prior on (-∞,∞)



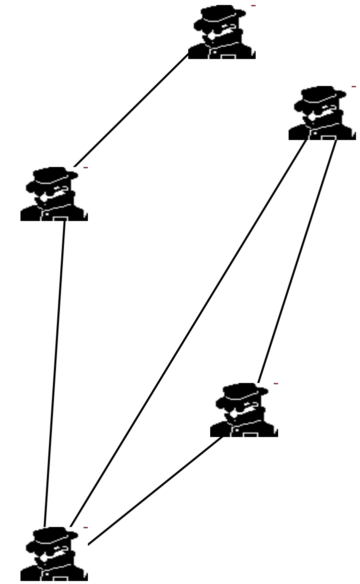
Results



- For every connected network:
- The best estimator is reached within $\mathbf{n}^2$ rounds where

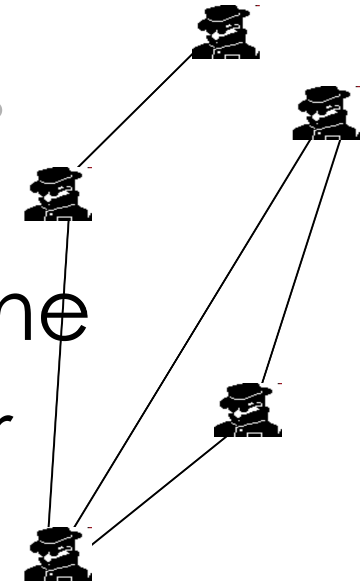
  n = #nodes (DVZ)
- Convergence time can be improved to 2* n * diameter (MT)
- All computations are efficient (MT)

Pf: ML and Min Variance.



- Claim 1: At each iteration
  $X_v(\mathbf{t})$ = Bayes Estimator
  = Maximum Like estimator
- Moreover, $X_v(\mathbf{t}) \in L_v(\mathbf{t})$, where
  $L_v(\mathbf{t})$= span { $X_w(\mathbf{0}),\ldots,X_w(\mathbf{t-1})$: w ~ v}
- $X_v(\mathbf{t})$ is a minimizer of
  $\{Var(X) : X \in L_v(\mathbf{t}), E[X] = \mu\}$
- Claim: Can be calculated efficiently

Pf: ML and Min Variance.



- Cor: $\mathrm{Var}(X_v(t))$ decreases with time
- Note: If $X_v(t) \neq X_u(t)$, dim of either L_v or L_u goes up by 1 ($v \sim u$)
- $\Rightarrow$ Converges in n^2 rounds.
- Claim: Weight that agent gives own estimator has to be at least 1/n (prove it!)
- $\Rightarrow$ converges to optimal estimator

Convergence in 2n*d steps

- Claim: If an agent u estimator X remains constant for 2*d steps t,t+1,…t+2d then the process has converged.

- Pf:
- Let $L = L_u(t+2d)$
- Let v be a neighbor of u.
- $X_{t+1}(v),\ldots X_{t+2d-1}(v) \in L$.
- $X \in L_v(t+1)$
- So $X_{t+1}(v) = \ldots = X_{t+2d-1}(v) = X$
- If w is a neighbor of w then:
- $X_{t+2}(v) = \ldots = X_{t+2d-2}(v) = X$
- By induction at time t+d all estimators are X.

- Open: Is there a bound that depends only on d?

Some open problems

• When is learning achieved? e.g. positively correlated signals?

•General statements about convergence times?

• More models where convergence time can be estimated?

•Efficiency of computations?

Truncated information

- Why could we analyze the cases so far?

- A main feature was that agents declarations were martingales.

- A more difficult case is where agents declarations are more limited.

- Example: +/- actions / declarations.

- This will be discussed next lectures.