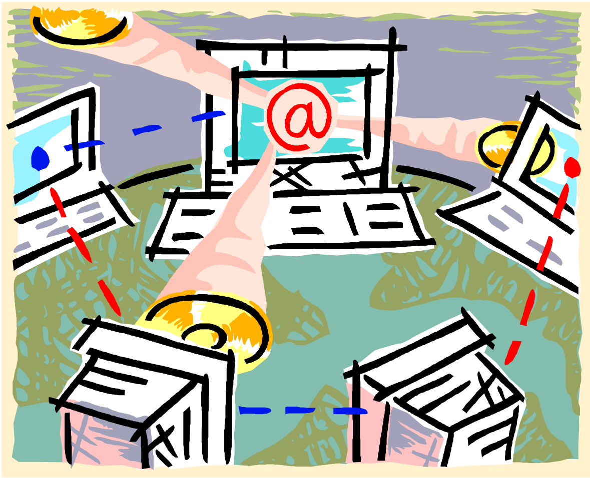
Social Choice and Social Networks

Bayesian Martingale Models

Elchanan Mossel
UC Berkeley



The Bayesian View of the Jury Theorem

- Recall: we assume +/- 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:
- $P[s = + \mid x] / P[s = - \mid x]$.

- Obtain posterior probability of +,-.

- Everybody agree about the posterior.

Critique of The Bayesian View

• The main critique is:

• In real elections people don’t all converge to the same posterior!

• The common prior assumption is obviously violated

• However, the Bayesian setup may still be useful:

Usefulness of the Bayesian View

- However, the Bayesian setup is still useful:

- Since it is has nice theory.

- It allows to compare different networks, modes of communication etc.

- Allows to test in what way people deviate from “rational behavior”

- Perhaps more applicable to learning: ask people to predict outcome of elections

- Perhaps more applicable to computational agents.

Challenges in The Bayesian View

- In Condorct Jury Theorem – the theory was easy.
- Why?
-

Challenges in The Bayesian View

- In Condorct Jury Theorem – the theory was easy.

- In general: the theory is easy if every agent can see the information of all other agents at some finite time.

- Theory is more interesting if only partial information is revealed. Examples:

- Each player only says how much she believes in something and not why.

- You only see some of the agents and not all.

A few examples of Bayesian Analysis

• In the first family of examples the goal is to evaluate the expected value of some function (prob. of some event).

• 2 players – agreeing to disagree (Aumann 1976)
• General directed graph (Parikh Krasucki 90s)
• Guassian signals (P. DeMarzo, D. Vayanos, and J. Zwiebel , M+Tamuz)

• In the 2nd family of examples the actions of players are very limited (binary) while the signal space is very rich (continuous).

• Voting on social networks (Gale Kariv 2003)
• The complete graph case (M + Tamuz)

Aumann's example

- Two agents have a complete common prior.
- Agent i=1,2 initially receives signal s(i).
- There is a bounded function f from the space to R say.
- Then for each time t:
- Agent 1 declares f(2t) = E[f | s(1),f(1),…,f(2t-1)]
- Agent 2 declares f(2t+1) = E[f | s(2),f(1),..,f(2t)]
- <u>Th (Aumann 76, Geanakoplos & Polemarchakis 82)</u>
- The sequence f(t) converges almost surely.

- Interpretation: let f be the indicator of some event.
- By repeatedly announcing their beliefs of the event the two agents will converge to the same posterior probability.
-
- Examples: Biased dice and samples.

Aumann's example

- Agent 1 declares f(2t) = E[f | s(1),f(1),…,f(2t-1)]
- Agent 2 declares f(2t+1) = E[f | s(2),f(1),..,f(2t)]
- Th (Aumann 76, Geanakoplos & Polemarchakis 82)
- The sequence f(t) converges almost surely.

- Proof idea
- Let F(t) denote the sigma algebra generated by the functions {f(1),…,f(t)}.
- Then E[f | F(t)], $t \geq 0$ is bounded martingale = view from the outside.
- Moreover: f(t) = E[f | F(t)] a.s. $\blacksquare$

- Comment: Note that the same argument applies to
- n agents as well.

A generalization to directed graphs

- We now consider the same story but with n agents on a directed graph G:
- At time t each vertex v declares it's expected value of f conditioned on its signal and what it has seen up to time t:
- $f(v,t) := E[f \mid s(v), f(w,s) \text{ , } w \in N(v), 1 \leq s \leq t-1]$
- Directed/Undirected <--> Phone vs. Email.
- Social Network aspect.
- Assume social network is known.

- Example: interval of length 3 and dice.

•<u>Q:</u> Do f(v,t) all converge to the same value?

A generalization to directed graphs

- Q: Is it the case that f(v,t) all converge to the same value?

- Obviously not:
- If there are two connected components they will not converge to the same value.
- In fact the graph u → v also does not converge.

-

A generalization to directed graphs

- Q: Is it the case that f(v,t) all converge to the same value?

- Obviously not:
- If there are two connected components they will not converge to the same value.
- In fact the graph u → v would also not converge.

- Thm (Parikh, Krasucki):
- In 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.

A generalization to directed graphs

Proof Sketch: :

- Let $F(v,t)$ be generated by $\{f(v,s) : s \leq t\}$ and conclude that $f(v,t)$ converges to $f(v) = E[f \mid F(v)]$, $F(v) = \{f(v,s)\}$
- $f(v)$ is the function closest in $L^2(F(v))$ to f.
- Next we do the same with $F'(v,t)$ generated by
- $\{f(v,s) : s \leq t\} \cup \{f(w,s) : s < t : w \in N(v)\}$
- Again we get that $f(v,t)$ converges to $f(v) = E[f \mid F'(v)]$
- Implies that if $v \rightarrow w$ in G then $|f(v)-f|_2 \leq |f(w)-f|_2$.
- Strongly connectivity $\Rightarrow \forall$ u,v: $|f(v)-f|_2 = |f(w)-f|_2$
- If $v \rightarrow w$ and $f(v) \neq f(w)$ then $g = 0.5(f(v)+f(w)) \in F'(v)$ and g closer to f than either $f(v)$ or $f(w)$.
- Strongly connectivity $\Rightarrow \forall$ u,v: $f(v) = f(w)$.

Some Things we do know about the model

- Players do not have to converge to the correct posterior.

- Example (Greg): prior (0.5,0.5) two players are given uniformly at random two bits whose e-xor is the state.

- For a finite state space: # of steps to convergence is at most # of sigma-algebras on the state.
- Pf: (Geanakoplos & Polemarchakis; Joe):
- When the two sigma-algebras remain the same for both players this will remain like that forever.

- More in GP: Examples where for n steps nothing happen and then converge to the same opinion.

Some Things we do know about the model

- Example: State space $[n^2]$ with uniform prior.
- Player 1 observes groups {1…,n},{n+1,..,2n} etc.
- Player 2 observes groups {1,…,n+1},…, n^2}
- True value is 1.
- The event is {1,n+2,2n+3,…, 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).
- etc.

Many things we do not know about this model

Many things we do not know about this model

• We do not know how long it takes to converge.
• We do not if it converges to a “good answer”.
• What is the computational complexity of the Bayesian process?
• It is known that if the original space is finite convergence will hold after finitely many steps.

.

Some aspects of the Bayesian approach

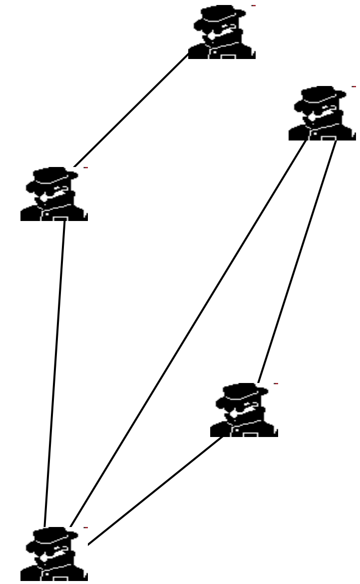
• We do not know how long it takes to converge.
• We do not if it converges to a “good answer”.
• What is the computational complexity of the Bayesian process?

•Some partial answers are known.

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

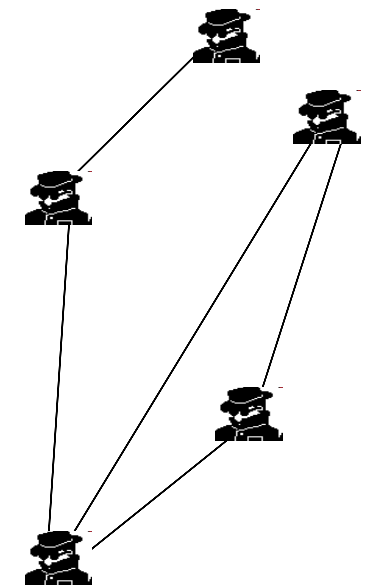
• Following model was studied in P. DeMarzo, D. Vayanos, and J. Zwiebel. and by Mossel and Tamuz.

The Gaussian Model



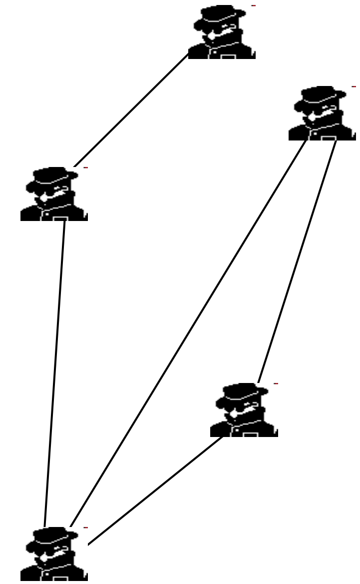
- The original signals are N(μ = ?,1).
- In each iteration
  - Each agent action reveals her current estimate of μ to her neighbors.
  - E.g. set price by min utility $(x- \mu)^2$
  - Each agent calculates a new estimate of μ based on her neighbors' broadcasts.
- Assume agents know the graph structure.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

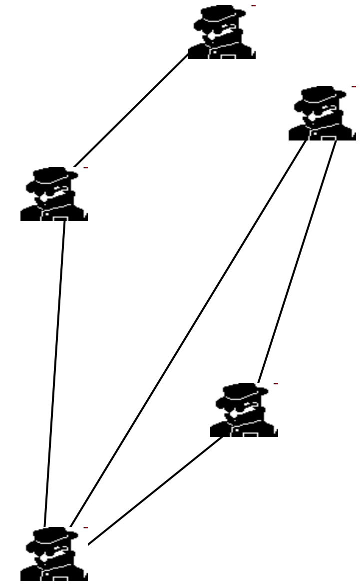
Results



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

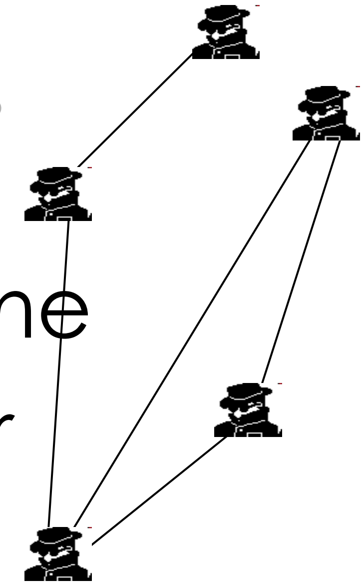
  n = #nodes (DVZ & MT)
- Convergence time can be improved to 2* n * diameter (MT)
- All computations are efficient (MT)

Pf: ML and Min Variance.



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

Pf: ML and Min Variance.



- Cor: $\mathbf{Var}(X_v(\mathbf{t}))$ decreases with time
- Note: If $X_v(\mathbf{t}) \neq X_u(\mathbf{t})$, dim of either $\mathbf{L}_v$ or $\mathbf{L}_u$ goes up by 1 ($v \sim u$)
- $\Rightarrow$ Converges in $\mathbf{n}^2$ 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 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 u then:
- $X_{t+2}(v) = \ldots = X_{t+2d-2}(v) = X$
- By induction at time t+d all estimators are X.

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 week.