Social Networks and Social Choice	**Lecture Date: September 14/16, 2010**

Bayesian Martingale Models

Lecturer: Elchanan Mossel *Scribe: N. M. Tran*

1 The Bayesian view of the Jury Theorem

1.1 Criticism on the Bayesian view

Recall the model setup: let $s \in \{\pm 1\}$ be the true state of the world. We shall assume a uniform prior on s, that is, $P(s = +) = P(s = -) = \frac{1}{2}$. Each voter receives signal $x_i \in \{\pm 1\}$, x_i are independent, and $P(x_i = s) = p$ for all i.

If this is indeed the case, then after all the vost has been cast, all voters can calculate

$$\frac{P(s = +|x)}{P(s = -|x)} = \frac{\prod_i P(x_i|s = +)}{\prod_i P(x_i|s = -)} = (\frac{p}{1-p})^{n_+ - n_-} =: r$$

where n_+ =number of voters who voted $+$, and n_- =number of voters who voted $-$.
From here, one can obtain the posterior distribution of s, which is

$$P(s = -|x) = 1 - P(s = +|x) = 1 - rP(s = -|x), \Rightarrow P(s = -|x) = \frac{1}{1+r}.$$

And now everybody agrees on the posterior. Since in real elections people don't all converge to the same posterior, this means the model is unrealistic: the common prior assumption is clearly unreasonable. However, the Bayesian setup may still be useful for the following reasons.

- It has a nice theory.
- It allows one to compare different networks, modes of communications,...
- Viewing the model as a standard for 'rational behavior', one can compare the results obtained from the model and reality to get a measure of how much people deviate from this 'rational behavior'.
- It can be applicable to learning, for example, in predicting outcome of elections.
- It can be applicable to computational agents (automated machines, where this 'rational behavior' may be a reasonable assumption.

1.2 Challenges in the Bayesian view

In general, the theory is easy if every agent can see the information of all other agents at some finite time, since everyone can then do the same posterior calculation, and so they all converge to the same decision in the next step. The theory is more interesting if the signal space is continuous (that is, $x_i \in \mathbb{R}$ as opposed to $\{\pm 1\}$), and/or only partial information is revealed. For example:

- Suppose $x_i \in \mathbb{R}$. Each player may only say how much she believes in something and not why. For example, each player only declares posterior value for p, but not her signal x_i.
- Suppose $x_i \in \mathbb{R}$. An agent only see some other agents but not all. This is leads to voting models on social networks, where agents are nodes on graphs and only see some form of information from her neighbors.

We still study the following two family of examples in details.

- The Aumann model [4, 5, 2, 1]. The goal is to evaluate the expected value of some function, for example, probability of some event (such as $P(s = +)$)
- The Gale-Kariv model. Here actions of players are binary (declare to believe or not), while the signal space is rich (is continuous since $x_i \in \mathbb{R}$).

2 Aumann's example

2.1 The 2-agent setup

Suppose that two agents have completely common prior. Agent i initially receives signals $s(i) \in \mathcal{S}$ for some signal space $\mathcal{S}$. Let $f : \mathcal{S} \to \mathbb{R}$ be a bounded function. Then at time t:

- Agent 1 declares $f(2t) = E(f|s(1), f(1), \ldots, f(2t-1))$
- Agent 2 declares $f(2t+1) = E(f|s(2), f(1), \ldots, f(2t))$.

Theorem 2.1. *[4, 2] The sequence $f(t)$ converges almost surely.*

Proof. Let $\mathcal{F}_t$ denote the sigma algebra generated by $\{f(1), \ldots, f(t)\}$. Then the sequence $\{f(t)\}$ is a martingale, and since f is bounded, it is uniformly integrable. Hence by the martingale convergence theorem, $\lim_{t\to\infty} f(t)$ exists almost surely. □

Suppose that f is the indicator of some event. Then by repeatedly announcing their beliefs of the event, the two agents will converge to the same posterior probability. However, the limit might **not** equal to $E(f|s(1), s(2))$, which is the Bayes posterior that both agents would reach had they given each other the signal instead of their declarations at each time step. Indeed, consider the following example.

Example 2.2. *Let $(S, X, Y) \in \{0, 1\}^3$, where S is the true state of the world, $X = s(1)$, $Y = s(2)$. Let (S, X, Y) be uniformly distributed on the set*

$$\{(x, y, z) : x + y + z = 0 \quad mod\ 2\}$$

Let $f = 1_{S=1}$. Then $E(f|X, Y) = 1_{X+Y=1}$ takes value either 0 or 1, while $f(1) = f(2) = \frac{1}{2}$. So the posterior of the two agents converge to $\frac{1}{2}$ after 1 step, but this is not the same as $E(f|X, Y)$.

2.2 The n-agent setup

Suppose we have n agents. If each agent can see the declaration of all other agents, then the proof for $n = 2$ generalizes directly and we have almost sure convergence. A more interesting generalization is one where each agent only knows the declaration of its neighbors:

Model setup: Let $G = (V, E)$ be a directed graph on n nodes, where there is an edge from i to j if j receives declarations of i. At time t, each vertex v declares its expected value of f conditioned on its signal and what it has seen up to time t. That is,

$$f(v, t) := E(f|s(v), f(w, s), w \in N(v), 1 \leq s \leq t - 1)$$

where $N(v) = \{w \in V, (w, v) \in E\}$, that is, $N(v)$ is the set of neighbors of v.
Directed/undirected graph can be thought of as modeling different modes of communications. Now, assuming that the social network is known, we want to know:

Question Do $f(v, t)$ all converge to the same value?

Clearly if the graph has disconnected components, or if a node as out-degree 0, then no convergence can occurs. What is interesting is that these are the only pathological behavior that can happen.

Theorem 2.3. *[5] If the graph G is strongly connected, then all agents will converge almost surely to the same value.*

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

Proof. Let $\mathcal{F}(v,t)$ be the sigma-algebra generated by $\{f(v,s) : s \leq t\}$, that is, the information known to agent v up until time t. Let $\mathcal{F}(v) := \lim_{t\to\infty} \mathcal{F}(v,t)$. For each fixed node v, $\{f(v,t) : t \geq 0\}$ is a martingale, with f bounded, hence by the martingale convergence theorem $f(v,t) \to f(v) = E(f|\mathcal{F}(v))$ almost surely. This implies $f(v)$ is the function closest to f in the set $L^2(\mathcal{F}(v))$ of square-integrable functions on $\mathcal{F}(v)$. In other words, we know that each agent converges almost surely. Now we just need to show that they all agree.

Now let $\mathcal{F}'(v,t)$ be the sigma-algebra generated by $\mathcal{F}(v,t) \cup \mathcal{F}(N(v),t)$ where $\mathcal{F}(N(v),t)$ is generated by $\{f(w,s) : s \leq t, w \in N(v)\}$. Let $\mathcal{F}'(v) := \lim_{t\to\infty} \mathcal{F}'(v,t)$. Again we see that $f(v,t)$ converges to $f(v) = E(f|\mathcal{F}'(v))$. This implies that if $(v,w) \in E$, then $\|f(v)-f\|_2 \leq \|f(w)-f\|_2$. By strong connectivity, this implies for all $u,v \in V$, $\|f(u)-f\|_2 = \|f(v)-f\|_2$. And this implies uniqueness of $f(v)$, for if $(v,w) \in E$ and $f(v) \neq f(w)$, then $g := 0.5(f(v)+f(w)) \in \mathcal{F}'(v)$ is closer to f than either $f(v)$ or $f(w)$, contradicting the fact that $f(v)$ is the L^2-minimizer to f with respect to the sigma algebra $\mathcal{F}'(v)$. Therefore we have $f(u) = f(v)$ for all $u,v \in V$. □

Again, as demonstrated in the $n = 2$ case, the players do not have to converge to the 'correct' posterior, that is, $E(f|s(v)\forall v \in V)$.

Theorem 2.4. *[2] If the state space is finite, then the number of steps to convergence is at most the number of sigma-algebras on the state (that is, the power set* $2^{|\Omega|}$*)*

Proof. ([2], Joe): Since after $2^{|\Omega|}$ steps, $\mathcal{F}(u,t) = \mathcal{F}(v,t)$ for all $u,v \in V$. As the sigma algebras remain unchanged in each iteration, the declaration of all players also remain unchanged forever. □

Note that $f(v,t) = f(v,t+1)$ for all v is **not** a sufficient indicator that $f(v)$ has converged. Here is an example in which the declaration of the two players are unchanged for n steps, then at the $n+1$ step they both converge to the same value.

Example 2.5. *Let the state space be* $[n^2] = \{1,\ldots,n^2\}$ *with uniform prior. Player 1 observes groups* $\{1,\ldots,n\},\{n+1,\ldots,2n\}$*, etc... Player 2 observes groups* $\{1,\ldots,n+1\},\ldots,\{n^2\}$*. Let* $f = 1_A$ *where* A *is the set* $\{1, n+2, 2n+3, \ldots, n^2\}$*. Suppose that the true value* ω *is 1, and that each player observes signals* $s = \{0,1\}^n$ *indicating which block* ω *falls in. (So in this case, for* $\omega = 1$*, both players get signals* $s(1) = s(2) = (1,0,\ldots,0)$*. Then we see that:*

- *At time 1:* $f(1) = \frac{1}{n}$*. Player 2 learns nothing from this declaration, so he declares* $f(2) = \frac{1}{n+1}$*.*
- *At time 2: Player 1 now learns that* $\omega \neq n^2$*. With this extra information, he still says* $f(3) = \frac{1}{n}$*. Player 2 now learns that* $\omega \notin$ *last group of player 1. Then we still have* $f(4) = \frac{1}{n+1}$

- *At time 3: player 1 now learns that $\omega \notin$ the two last groups of player 2. Declare $f(5) = \frac{1}{n}$. Similarly, player 2 now learns that $\omega \notin$ the two last groups of player 1. Declare $f(6) = \frac{1}{n+1}$.*

The process goes on, until in the n^{th} step, which both players know that $\omega \in [n] \cap [n+1]$. Then both declare $f(t) = \frac{1}{n}$, and this is the value of f at the limit.

2.3 Open problems on the Aumann family of models

A number of properties of these models remain unknown. For example:

- What is an optimal bound on the number of iteration to convergence in terms of the state space?
- Under what conditions would it converge to a 'good answer' (ie: same as the Bayes posterior $E(f|$ all signals)?
- What is the computational complexity of the Bayesian process?

For certain specific subclasses of models, some of these answers are known. We will study a Gaussian model which is computationally feasible, has rapid convergence, and converges to the optimal answer for every connected network.

2.4 The Gaussian Model

This model was studied in [3, 1]. The setup is a special case of the n-agent model:

- $\mathcal{S} = \mathbb{R}$, signals $s(i) \sim N(\mu, 1)$ for some unknown mean μ.
- At each iteration, each agent reveals her current estimate of μ to her neighbors. In the notation of the n-agent model, $f(v,t)$ is some estimator of μ
- Each agent calculates a new estimate of μ based on her neighbor's broadcasts.
- Assume that all agents know the graph structure, and the process goes on for infinitely many iterations.

Note that given all the signals $s(1), \ldots, s(n)$, then the estimator $\overline{s} := \frac{1}{n}\sum_{i=1}^{n} s(i)$ is optimal in the sense that it is both the MLE and the Bayes estimator with uniform prior on $\mathbb{R}$. Trivially, this estimator could be achieved by the all the agents if the graph is complete (everyone is friend with everyone else). In fact, it could be achieved for every connected graph.

Theorem 2.6. *For every conencted network, the best estimator $\overline{s}$ is reached within n^2 rounds, where n is the number of nodes in the graph. [3, 1]*
In fact, the number of steps it takes to converge is at most $2nd$ where d is the diameter of the graph. [1]
The computations (updating the value at each node at each time step) can be done efficiently. [1]

Proof. At each iteration, let $X(v,t) = f(v,t)$ denotes the maximum likelihood estimator for μ of agent v. Note that $X(v,t) \in L(v,t)$ where $L(v,t) = \text{span}\{X(w,0),\dots,X(w,t-1)\}$ where w is a neighbor of v. Therefore, $X(v,t) = \text{argmin}_X\{Var(X) : X \in L(v,t), E(X) = \mu\}$. In other words, it is the best unbiased estimator for μ in the space $L(v,t)$. This just comes from the fact that maximum likelihood estimator for μ in a normal distribution is the unbiased estimator with minimal variance.

Proof of convergence in n^2: Since $L(v,t) \subseteq L(v,t+1)$, $Var(X(v,t)) \geq Var(X(v,t+1))$, that is, the variance of $X(v,t)$ decreases with time. Now, if v and u are neighbors, and $X(v,t) \neq X(u,t)$, then the dimension of either $L(v,t)$ or $L(u,t)$ goes up by 1. Since the dimension of $L(v,t)$ is at most n, this implies that convergence takes place in n^2 rounds.

Proof of convergence to the optimal $\overline{s}$: Now we need to show that the convergence is to the optimal estimator. Note that we can write $X(v,t)$ as a weighted linear combination of the signal $s(v)$ that agent v receives, and signals $s(w)$ of her neighbors:

$$X(v,t) = aX(v,0) + bZ$$

where $X(v,0) = s(v)$, $Z \in \text{span}\{X(w,0)\}$ where w is a neighbor of v, and $a,b \in \mathbb{R}$ are weights. Note that $Var(Z) \geq \frac{1}{n-1}$, since $\frac{1}{n-1}$ is the variance of the estimator $\frac{1}{n-1}\sum_{i\neq v} s(i)$, which is the best estimator based on all signals except $s(v)$. Hence $Var(X(v,t)) \geq a^2 + \frac{b^2}{n-1}$. Since $a+b=1$, rearranging and solve for a, we see that $a \geq \frac{1}{n}$. In other words, the weight that each agent v gives her own estimator has to be at least $\frac{1}{n}$. Since all agents converge to the same estimator, and since the estimator is unbiased, the sum of the weights in the estimator must be 1. Hence the weights must be $\frac{1}{n}$, so $X(v,t)$ converges to the optimal estimator $\overline{s}$.

Proof of convergence in $2nd$ steps: This is equivalent to showing that if the estimator $X(u,t)$ of agent u remains the same for $2d$ steps, then the process has converged. To prove this, let $L := L(u,t+2d)$, let $X := X(u,t+2d)$ denote the unchanging estimator at u, and let v be a neighbor of u. Since $X(v,t+1),\dots,X(v,t+2d-1) \in L$, and since $X \in L(v,t+1)$, we have $X(v,t+1) = \ldots = X(v,t+2-1) = X$. If w is a neighbor of v, then $X(w,t+2) = \ldots = X(w,t+2d-2) = X$. Hence by induction, at time $t+d$ all estimators are X. $\square$

To have a better feeling for this process, one can consider an abstract example of 4 agents on a path.

Example 2.7. *4 agents on a path.*
Label the nodes from left to right by 1,2,3,4. Suppose at time 0*, they receive signals* a, b, c, d*.*

- *Time 1: the agents exchange the signals. Based on that, the updates are:*
  - $X(1,1) = \frac{a+b}{2}$
  - $X(2,1) = \frac{a+b+c}{3}$
  - $X(3,1) = \frac{b+c+d}{3}$
  - $X(4,1) = \frac{c+d}{2}$
- *Time 2:*
  - *Agent 1 knows that agent 2 has more information than she does, so from her viewpoint agent 2's estimator is optimal. Hence she will just copy the estimator of agent 2:* $X(1,2) = \frac{a+b+c}{3}$*.*
  - *Similarly, agent 4 will just copy the estimator of agent 3:* $X(4,2) = \frac{b+c+d}{3}$*.*
  - *Agent 2 knows the values a,b,c, and now she knows* $\frac{b+c+d}{3}$*. From here she can solve to get d, so now she states the optimal value:* $X(2,2) = \frac{a+b+c+d}{4}$*.*
  - *Similarly, agent 3 states the optimal value* $X(3,2) = \frac{a+b+c+d}{4}$
- *Time 3: agent 1 and 4 copies 2 and 3, hence converges. So the process converges in 3 steps.*

2.5 Open problems and comments

There are some open problems left in the Gaussian model case. For instance, the optimal bound on the number of steps it takes to converge is still unknown. [1] conjectured that this is of order $O(n)$, or possibly $O(n/d^*)$, where d^* is the minimal degree of the graph.

Note that the main feature for all the models we analyzed so far is that agent declarations are martingales. If the agents declarations are more limited, for example, of $\pm$ type, then the situation is more difficult. This will be studied in the next lecture, on the Gale-Kariv model.

References

[1] O. Tamuz E. Mossel. Efficient bayesian learning in social networks with gaussian estimators. 2010.

[2] H.M. Polemarchakis J.D. Geanakoplos. We can't disagree forever. 1982.

[3] J. Zwiebel P. DeMarzo, D. Vayanos. Persuasion bias, social influence, and unidimensional opinions. 2003.

[4] R. J. Aumann. Agreeing to disagree. 1976.

[5] P. Krasucki R. Parikh. Communication, concensus and knowledge. 1990.