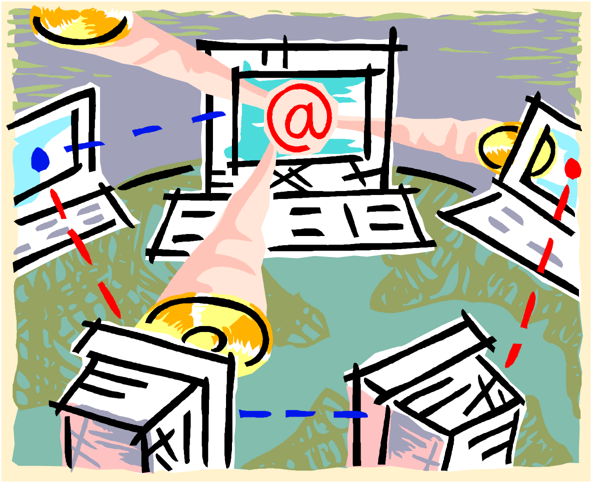
Probability Models of Information Exchange on Networks

Lecture 1

Elchanan Mossel
UC Berkeley



Motivating Questions

- How are collective decisions made by:
- people / computational agents

- Examples: voting, pricing, measuring.

- Biased / Unbiased private signals.

- Network structure.

- Types of signals (numbers, binary, behaviors etc.)

- Opinion leaders, communities.

Main Theme : Aggregation of Biased Signals

Condorcet's jury theorem:
"*Essay on the Application of Analysis to the Probability of Majority Decisions*, 1785":

• Juries reach a decision by majority vote of n juries.
• One of the two outcomes of the vote is *correct*, and
• Each juror votes correctly independently with probability $p>1/2$.

•Then in the limit as the group size goes to infinity, the probability that the majority vote is correct approaches 1.

• Major question: what can be done if communication between individuals is limited?

Main Theme : Aggregation of Biased Signals

Condorcet's jury theorem

• Major question: what can be done if communication between individuals is limited?

• Communication can be limited due to network effects (not all individuals see all others).

• Communication can be limited due to bandwidth (instead of telling me the distribution of temperature tomorrow, you just tell me it will be hot).

Lecture Plan

- Lecture 1: Condorcet's Thm in general setups.

- Lecture 2: Networks models, Markov chains and voter model.

- Lecture 3: Bayesian Network Beliefs Models.

- Lecture 4-5: Bayesian Network Actions Models.

- Lecture 6: Some other examples of information exchange: competition and marketing.

Condorect's Jury Theorem (1785)

• n juries will take a majority vote between two alternatives - and +.

• Either - or + is *correct* (prior 0.5 each)

•Each jury votes correctly independently w.p p > ½.

Then

• P[correct outcome] $\rightarrow$ 1 as n $\rightarrow \infty$
• This is referred to as "Aggregation of Information".

•Follows from LLN (Bernoulli 1713)

Condorcet (wiki)

Bernoulli (wiki)

Cardano (wiki)

The Condorcet Thm

• Condorcet proved more: P[correct] increases with n.

•Pf: Exercise. Also follows from:

• Neyman-Pearson Lemma (1933): Among all possible decision procedures, given the individual votes, the probability of correct estimation is maximized by a majority vote.

• Exercise: show that standard statement of NP implies this.

• Question: What if information exchange is limited?

Neyman (wikipedia)

Pearson (learn math)

How small can p be as a function of n for the conclusion to hold?

How small can p be as a function of n for the conclusion to hold?

- Recall the Central Limit Theorem.

How small can p be as a function of n for the conclusion to hold?

- Let $p(n) = 0.5 + c\, n^{-1/2}$ and
- $q(n,c)$ = P[Maj is correct] give n ind. $p(n)$ signals
- Then by the CLT

- $\lim q(n,c)$ is about $P(N(0,1) > -2c)$

- So if $p-0.5 \gg n^{-1/2}$ then $q(n) \to 1$.

- If $p-0.5 \ll n^{-1/2}$ then $q(n) \to 1/2$

- Exercise: Explain the conclusion!

- Exercise: calculate $\lim_n q(n,c)$

Moivre (wikipedia)

Laplace (wikipedia)

Beyond Condorcet's Jury Theorem

- Further questions:

- What about other aggregation functions?

- We will be mostly interested in functions obtained by agents interactions.

- But today some simpler examples.

The Electoral College example

• Assume $n = m^2$ = odd square.

• Consider an imaginary country partitioned into m states each with m voters.

• Consider the following voting rule:
• Winner in each state chosen according to majority vote in that state.
• Overall winner = winner in the majority of states.

• Questions:
• Is this method different than majority vote?
• Does the conclusion of the jury theorem still hold?

The Electoral College example

Questions:

- Is this method different than majority vote?

- Yes (for all $m > 2$).

- Does the conclusion of the jury theorem still hold?

- It does – by simple recursion.

$\blacksquare$

Small Bias in Electoral College

- Assume $n = m^2$ is an odd square.

- What is the smallest bias that guarantees the conclusions of the jury theorem?

Small Bias in Electoral College

- Assume $n = m^2$ is an odd square.

- What is the smallest bias that guarantees the conclusions of the jury theorem?

- Claim: Let $p = 0.5 + a/m = 0.5 + a\, n^{-1/2}$ and let
- $p(a)$ = probability outcome is correct as $m \to \infty$.
- Then: $p(a)$ is well defined and $p(a) \to 1$ as $a \to \infty$.

- Exercise: Prove this.
- Exercise: Write a formula for $p(a)$.
- Exercise: Let $q(a)$ be the corresponding quantity for majority. Show that $p(a) \leq q(a)$. Can you prove this directly?

More examples

- We can similarly try to analyze many more examples.

- Exercise: Compare Majority and electoral college in the US. What value of p is needed to get the correct outcome with probability 0.9? 0.99?

- To some general examples.

General functions

- What are the best/worst functions for aggregation of information?

- An aggregation function is just a function $\{-,+\}^n \rightarrow \{-,+\}$

Some bad examples

• What are the best/worst functions for aggregation of information?

• An aggregation function is just a function $\{-,+\}^n \rightarrow \{-,+\}$

• Answer:
• The function that does the opposite of Majority function doesn’t aggregate very well …

Monotonicty

- What are the best/worst functions for aggregation of information?

- The function that does the opposite of Majority function doesn't aggregate very well …

- This function is not natural. It is natural to look at monotone functions:

- f is monotone if $\forall\, i\ x_i \leq y_i \Rightarrow f(x) \leq f(y)$

- Q: What are the best/worst monotone aggregation functions?

An example

Q: What are the best/worst monotone aggregation functions?

• The constant (monotone) function f = + doesn’t aggregate very well either.

Fairness

Q: What are the best/worst monotone aggregation functions?

- The constant (monotone) function f = + doesn't aggregate very well either.

- We want to require that f is fair – treats the two alternatives in the same manner.

- f is fair if f(-x) = -f(x).

- Q: assuming f is monotone and fair what is f(++++++)?
-
- Q: What are the best/worst fair monotone aggregation functions?

Formal Statement

Q: What are the best/worst monotone aggregation functions?

- Recall:
- Apriori correct signal is +/- w.p. ½.
- Each voter receives the correct signal with probability $p > ½$.
- For a fair aggregation function f, let

$C(p,f)$ = P[f results in the correct outcome]

$= P[f = + \mid \text{signal} = +]$

Q: "What are the best/worst fair monotone aggregation functions?" means

Q: What are the fair monotone aggregation functions which minimize/maximize $C(p,f)$?

The Best Function

Claim: Majority is the best fair monotone symmetric aggregation function (not clear who proved this first – proved in many area independently)

Pf: Follows from Neyman Pearson

The Best Function

Claim: Majority is the best fair monotone symmetric aggregation function (not clear who proved this first – proved in many area independently)

Alternative pf: $C(f,p) = \sum_x P[x]\, P[f(x) = s \mid x]$

To maximize this over all f need to choose f so that $f(x)$ has the same sign as $(P[s = + \mid x] - P[s = - \mid x])$.

Now by Bayes rule:

$P[s = + \mid x] \,/\, P[s = - \mid x] = P[x \mid s=+] \,/\, P[x \mid s=-] =$

$= a$ to the power $\{\#(+,x)-\#(-,x)\}$

where $a = p/(1-p) > 1$ and $\#(+,x)$ is number of +'s in x

So optimal rule chooses $f(x) = \mathrm{sign}(\, n(+,x)-n(-,x))$

The Worst Function

Claim: The worst function is the dictator $f(x) = x_i$.

For the proof we'll use Russo's formula:

Claim 1: If f is a monotone function $f : \{-,+\}^n \to \{-,+\}$ and
$f_i(x) = f(x_1,\ldots,x_{i-1},+1,x_{i+1},\ldots,x_n) - f(x_1,\ldots,x_{i-1},-1,x_{i+1},\ldots,x_n)$

then $C'(f,p) = 0.5 \sum_{i=1}^n E_p[f_i] = \sum_{i=1}^n E_p[Var_{i,p}[f]]/(4p(1-p))$

$Var_i[f] = E_p[\, Var_p[f \mid x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n]\,]$

Pf: Use the chain rule and take partial derivatives.

Remark: f_i is closely related to the notion of pivotal voters (economics) and influences in computer science.

The Worst Function

Claim: The worst function is the dictator $f(x) = x_i$.

The second claim we need has to do with local vs. global variances:

Claim 2: $\mathrm{Var}[f] \leq \sum_i \mathrm{Var}_i[f]$ with equality only for functions of one coordinate.

Pf of Claim 2: Possible proofs:
Decomposition of variance of martingales differences
Fourier analysis

The Worst Function

Claim: The worst function is the dictator $f(x) = x_i$.

Claim 1: $C'(f,p) = \sum_{i=1}^{n} p\, E_p[f_i] = (2(1-p))^{-1} \sum i=1^n n\, Var_i[f]$

Claim 2: $Var[f] \leq \sum_i Var_i[f]$

Pf of main claim:

- For all monotone fair functions we have $C(g,0.5)=0.5$ and $C(g,1)=1$.
- Let f be a dictator and assume by contradiction that
- $C(f,p) > C(g,p)$ for some $p>1/2$.
- Let $q = \inf \{p : C(f,p) > C(g,p)\}$ then
- $C(f,q) = C(g,q)$ and $C'(f,q) \geq C'(g,q)$ so:
- $Var_q[g] = Var_q[f] = \sum_i Var_{i,p}[f] \geq \sum_i Var_{i,p}[g]$
- So g is function of one coordinate.

Other functions?

So far we know that:

1. Majority is the best.
2. Electoral college aggregates well.
3. Dictator is the worst among fair monotone functions and doesn’t aggregate well.
4. What about other functions?

5. Example: Recursive majority (todo: add details and pic)

6. Example: An extensive forum (todo: add details and pic).

The effect of a voter

<u>Def</u>: $E_p[f_i]$ is called the influence of voter i.

<u>Theorem (Talagrand 94):</u>

- Let f be a monotone function.
- If $\delta = \max_x \max_i E_x[f_i]$ and $p < q$ then
- $E_p[f \mid s = +] (1-E_q[f \mid s=+]) \leq \exp(c \ln \delta \, (q-p))$
- for some fixed constant c>0.

- In particular: if f is fair and monotone, taking p=0.5:

- $E_q[f \text{ is correct}] \geq 1- \exp(c \ln \delta \, (q-0.5))$

The effect of a voter

. <u>Theorem (Talagrand 94):</u>

- Let f be a monotone function.
- If $\delta = \max_p \max_i E_x[f_i]$ and $p < q$ then
- $E_p[f \mid s = +]\,(1-E_q[f \mid s = +]) \leq \exp(c \ln \delta\,(q-p))$
- for some fixed constant c>0.

- In particular: if f is fair and monotone, taking p=0.5:

- $E_q[f \text{ is correct}] \geq 1- \exp(c \ln \delta\,(q-0.5))$

- This means that if each voter has a small influence then the function aggregates well!

An important case

Def: A function $f: \{-,+\}^n \rightarrow \{-,+\}$ is transitive if there exists a
- group G acting transitively on $[n]$ s.t.
- for every $x \in \{-,+\}^n$ and any $\sigma \in G$ it holds that $f(x_\sigma) = f(x)$,

where
- $x_\sigma(i) = x(\sigma(i))$

Thm (Friedgut-Kalai-96) :
- If f is transitive and monotone and
- $E_p[f \mid s= +] > \varepsilon$ then
- $E_q[f \mid s = +] > 1-\varepsilon$ for $q=p+c \log(1/2\varepsilon)/ \log n$

Note: If f is fair transitive and monotone
we obtain
$E_q[f \text{ is correct}] > 1-\varepsilon$ for $q=0.5+c \log(1/2\varepsilon)/ \log n$

An important case

<u>Thm (Friedgut-Kalai-96) :</u>

- If f is transitive and monotone and
- $E_p[f] > \varepsilon$ then
- $E_q[f] > 1-\varepsilon$ for $q=p+c \log(1/2\varepsilon)/ \log n$

- <u>Note:</u> If f is fair transitive and monotone we obtain $E_q[\text{f is correct}] > 1-\varepsilon$ for $q=0.5+c \log(1/2\varepsilon)/ \log n$

- This implies aggregation of information as long as the signals have correlation at least $0.5+c/\log n$ with the true state of the world.

Examples of aggregation / no aggregation

Claim:

Examples: Electoral college

Example: Recursive Majority

Example: Hex Vote

Note: The results actually hold as long as there are finitely many types all of linear size in n.

Other distributions

So far we have discussed situations were signals were independent. What is signals are dependent?

Setup: Each voter receives the correct signal with probability p
But: signals may be dependent.

Question: Does Condorcet Jury theorem still hold?

Other distributions

So far we have discussed situations were signals were independent. What is signals are dependent?

Setup: Each voter receives the correct signal with probability p
But: signals may be dependent.

Question: Does Condorcet Jury theorem still hold?

A: No. Assume:

1. With probability 0.9 all voters receive the correct signal.
2. With probability 0.1 all voters receive the incorrect signal.

Other distributions

This example is a little unnatural. Note that in this case just looking at one voter we know the outcome of the election.

Other distributions

This example is a little unnatural. Note that in this case just looking at one voter we know the outcome of the election.

Def: The effect of voter i on function $f: \{0,1\}^n \rightarrow \{0,1\}$ for a probability distribution P is:

$e_i(f,P) = E[f \mid X_i = 1] - E[f \mid X_i = 0]$.

Note: Assume $E[X_i] = p$ then:

$$\begin{aligned} Cov[f,X_i] &= E[f*(X_i - p)] = \\ &= p\, E[(1-p)\, f \mid X_i = 1] + (1-p)\, E[-p\, f \mid X_i = 0] \\ &= p(1-p)\, e_i(f,P) \end{aligned}$$

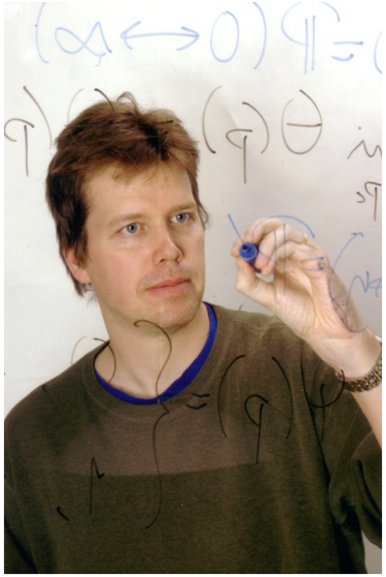
Condorcet's theorem for small effect functions

<u>Theorem (Haggstrom, Kalai, Mossel 04):</u>

• Assume n individuals receive a 1,0 signal so that $P[X_i = 1] \geq p > ½$ for all i.

• Let f be the majority function and assume $e_i(f,P) \leq e$ for all i.

• Then P[f is correct] > 1 – e/(p-0.5).



Condorcet's theorem for small effect functions

<u>Theorem (Haggstrom, Kalai, Mossel 04):</u>

- Assume n individuals receive a 1,0 signal so that $P[X_i = 1] = p_i \geq p > ½$ for all i.
- Let f be the majority function and assume $e_i(f,P) \leq e$ for all i.
- Then P[f is correct] > 1 – e/(p-0.5).
-
- Proof: quite easy …

Condorcet's theorem for small effect functions

<u>Theorem (Haggstrom, Kalai, Mossel 04):</u>

- Assume n individuals receive a 1,0 signal so that $P[X_i = 1] = p_i \geq p > ½$ for all i.
- Let f be the majority function and assume $e_i(f,P) \leq e$ for all i.
- Then P[f is correct] > 1 – e/(p-0.5).

- Proof: Let $Y_i = p_i - X_i$ and $g= 1- f$. Then

- $E[(\sum Y_i) g] = E[g]\, E[\sum Y_i \mid f = 0] \geq n\,(p-1/2)\, E[g]$

- $E[(\sum Y_i) g] = \sum E[Y_i\, g] = \sum \mathrm{Cov}[X_i,f] = \sum p_i(1-p_i)\, e_i(f) \leq n\, p(1-p)\, e$

So $n(p-1/2)E[g] \leq n\, p\,(1-p)\, e \Rightarrow$

$E[g] \leq ep(1-p)/(p-0.5)$

$E[f] \geq 1 – ep(1-p)/(p-0.5)$.

Comments about the proof

- Proof actually works for all weighted majority functions.

- So for weighted majority functions we have aggregation of information as long as they have small effects.

- In fact the following is true:

<u>Theorem (HKM-04)</u>
- If f is transitive, monotone and fair and is not simple majority then there exists a probability distribution so that:
$E[X_i] > ½$ for all i and $E[f] = 0$ and $e_i(f,P) = 0$ for all i.

- If f is monotone and fair and is not simple majority then there exists a probability distribution so that:
$E[X_i] > ½$ for all i and $E[f] = 0$ and $e_i(f,P) = 0$ for all i.

Comments about the proof

•

Theorem (HKM-04)

• If f is transitive, monotone and fair and is not simple majority then there exists a probability distribution so that:

$E[X_i] > ½$ for all i and $E[f] = 0$ and $e_i(f,P) = 0$ for all i.

Philosophical Perspective

- Egalitarian voting systems are "better".

- Condorcet: Bigger Majorities are better.

- Neyman Pearson: Majority is best for independent signals.

- Dictator is the worst.

- For general dependent measures, majority is the only transitive voting methods that aggregates small effects voters.

Next Lectures

- Analyze natural dynamics of aggregation and check how well they aggregate.

- Egalitarian systems will often be better.

- In general it is hard to “check” if a certain voting system has small effects or not.

- Some of the natural voting methods won’t be monotone!

Additional Exercise 1

1 Suppose $X_1,\ldots,X_n$ are ind. Signals which are correct with probabilities $p_1,\ldots,p_n$. And $Y_1,\ldots,Y_n$ are ind. Signals which are correct with probability $q_1,\ldots,q_n$.

-
- Assume that f is monotone and fair and that it returns the correct signal for the X's with probability at least $1-\delta$. Show that the same is true for the Y's if $q_i \geq p_i$ for all i.

- In words – if f aggregates well under some signals it aggregates even better under a stronger signal.

HW2

- Consider the electoral college example with m states of size m each where m is odd.

- Show that a signal of strength 0.5 + 1000/m results in an aggregation function which returns the correct result with probability at least 0.99 for all m sufficiently large.

- Hint: Use the local central limit theorem.

Additional Exercise 2

- Compare the actual electoral college used in the US in the last elections to a simple majority vote in terms of the quality of independent signals needed to return the correct result with probability 0.9 and 0.99.

Additional Exercise 3

- Let $f : \{-,+\}^n \rightarrow \{-,+\}$. Consider i.i.d. $X_1,\ldots,X_n$ such that $P[X_i = +] = p$. Show that $Var[f] \leq \sum Var_i[f]$.