# Stochastic Consensus and the Shadow of Doubt

Emilien Macault, CREST & BETA Economics

Blockchain@X-OMI , Sept. 21st, 2023

# Motivation

We study **opinion dynamics** in a **network of agents**: how do opinions evolve over time?

Core question in economic theory: **opinion aggregation** started with Condorcet's jury (1785).

Applications in consensus estimation, optimal belief-manipulation, preventing disinformation, estimating the impact of lobbying strategies, ...

**Stylized facts:** increasing network density and scale, unverifiable information, large spread of marginal opinions, limits to rationality.

# Summary

Contributions of this paper:

- We propose a new model of non-Bayesian belief exchange;
- We prove that under general conditions, beliefs converge;
- We characterize conditions that ensure a consensus emerges;
- We show that in general, this consensus is a full-support random variable;
- We provide characterization elements for this limit belief.

Next step: testing in the lab!

# Stochastic Opinion Formation

Key assumption: communication is **stochastic**. Agents may spread marginal beliefs as long as they put positive probability on them.

Agents communicate by iteratively drawing possible states of the world according to their beliefs $\rightarrow$ no metacognition.

We use **reinforcement learning** techniques by modeling the opinion dynamics as an **interacting urn system**.

Our goal $\rightarrow$ study the **dynamics of beliefs**: convergence, consensus, nature of the limit beliefs.

# Literature

**Non-Bayesian/boundedly rational opinion formation**: DeGroot (1974), Friedkin & Johnsen (1990, 1997), Bala & Goyal (1998), De Marzo et al. (2003), Golub & Jackson (2007, 2010), Jadbabaie et al. (2012), Venel & Mandel (2020);

**Robustness**: Golub & Jackson (2010), Acemoglu & Ozdaglar (2011), Acemoglu et al. (2016), Golub & Sadler (2017), Peretz et al. (2021);

**Iteracting urn models**: Eggenberger & Polya (1923), Hill et al. (1987), Paganoni & Secchi (2004), Dai Pra et al. (2014), Crimaldi et al. (2016), Lehrer & Shaiderman (2018);

**Stochastic methods**: Monro & Robbins (1951), Benaïm (1999), Renlund (2000), Kushner & Yin (2003), Pemantle (2007), Borkar (2009), Laruelle et al. (2013)

**Experimental/Simulations**: Molavi et al. (2017), Chandrasekhar et al. (2020), Banerjee et al. (2021), DeFilippis et al. (2022), Molavi (2022).

# Model Premises

Our model is defined by:

- A finite set of agents $V = \{1, \ldots, N\}$;
- A fixed, undirected and non-weighted communication network $G = (V, E)$;
- A binary state of the world $\theta$;
- Each agent holds a prior belief on $\theta$ (cf. next slide).

Contrary to Condorcet, the existence of a "true" value for $\theta$ is not important: we focus on beliefs dynamics.

We consider pure informational dynamics $\rightarrow$ this is not a game *per se*, no payoffs, no strategies.

# Beliefs and Reinforcement Learning

Each agent's belief is modeled as an **urn** containing red and blue balls representing the two possible values of $\theta$.

**Condorcet initialization**: at $t = 0$ one red w.p. $\alpha$ and one blue w.p. $1 - \alpha$

Most results hold independently of prior beliefs $\rightarrow$ $\alpha$ mostly serve comparative statics.

No additional information and/or feedback after the initialization.

# Communication process

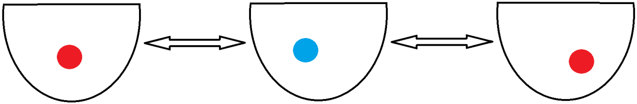
Communication proceeds in discrete time $\rightarrow$ at each time $t \geq 1$:

1. Every agent draws a ball from her urn with replacement.
2. Agents observe their neighbors draws.
3. Agents reinforce their beliefs: add one ball for each observed draw.

Beliefs dynamics: do they converge? Do they coincide in the limit? Can we say anything about this limit?

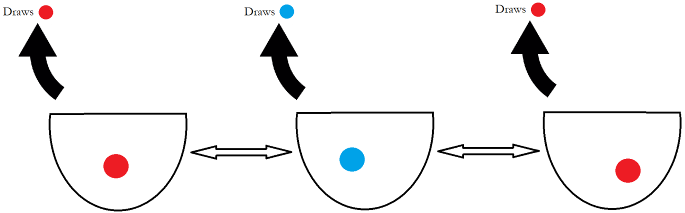
# Example: $t = 1$

Consider 3 agents connected in a line with the following initial compositions. At time $t = 1$, we have:



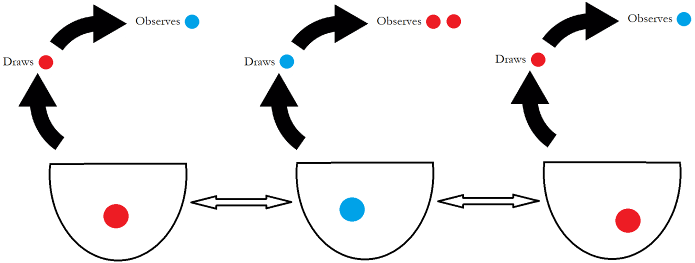
# Example: $t = 1$

Consider 3 agents connected in a line with the following initial compositions. At time $t = 1$, we have:



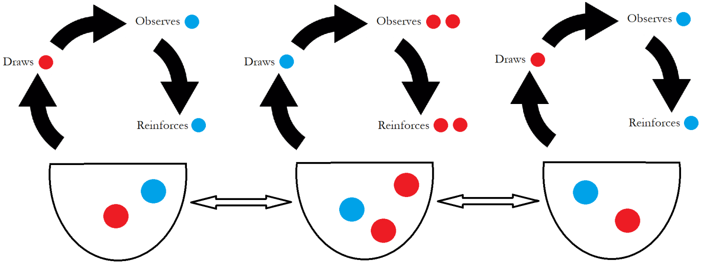
# Example: $t = 1$

Consider 3 agents connected in a line with the following initial compositions. At time $t = 1$, we have:



# Example: $t = 1$

Consider 3 agents connected in a line with the following initial compositions. At time $t = 1$, we have:



# Convergence

We first prove that **beliefs converge almost-surely**.

**Theorem**

*For every communication network and every initial composition of urns, proportions in each urn converge almost-surely.*

*Why?* Stochastic approximation: over time, beliefs approach trajectories of a continuous-time ODE. In the limit, proportions converge with probability 1 to a bounded invariant set of the ODE.

# Consensus

We show a **consensus emerges** in **connected components**.

Theorem

*For every initial composition of urns, if the communication network is connected, then the limit proportions in every urn are equal almost-surely.*

*Why?* Spectral properties of the network's Laplacian matrix. We characterize the unique stable invariant set of the ODE as the set of vectors with identical coordinates.

Conversely, limit proportions within each subnetwork are all equal.

# Full support

The limit belief is not deterministic but a **full-support random variable**:

**Theorem**

*For any communication network and any* $\alpha \in (0,1)$*, the limit belief is a non-trivial random variable with full support over* $[0,1]$*.*

*Why?* Any invariant point $p$ of the ODE is *attainable* in finite time with positive probability. We then show that $p$ must belong to the support of the limit distribution.

# A few comments

The average proportion of a color in the limit belief is increasing with the number of balls of said color in the initial composition of urns (more on that in a few slides).

Yet, one single ball of a given color in a single urn can overturn the limit belief: with at least one red ball, for any $x \in (0, 1)$, the probability that the limit proportion of red is greater than $x$ is strictly positive.

Key difference with predictions from usual models: we do not converge to a point anymore. It leaves room for butterfly effect.

# Beta distribution

We simulate the process on star, complete and $k$-regular networks and fit candidate distributions.

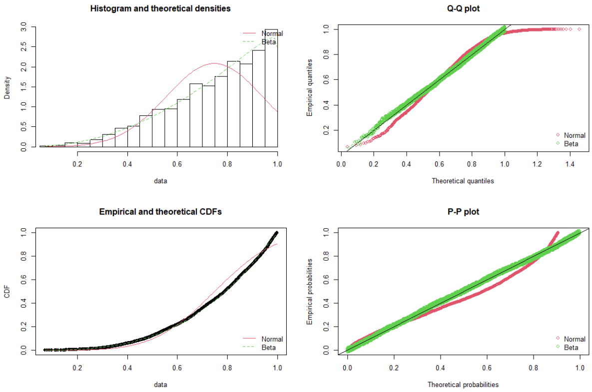



Figure: Beta and normal fits on a star network with $\alpha = 0.75$, 5000 obs.

# Average Limit Belief

We observe that the empirical average of the limit belief is equal to $\alpha$ for every network topology.

**Theorem**

*For any value of $\alpha$, if the communication network is regular, then the conditional expectation of the limit belief ex-ante is equal to $\alpha$, proof for the regular case.*

| $\alpha$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|
| Av. | 0.097 | 0.20 | 0.29 | 0.40 | 0.49 | 0.60 | 0.69 | 0.79 | 0.89 |

Table: Empirical mean in a star network of size $N = 50$ (7500 obs.).

On average, proportions remain the same. But the variance is non-zero and depends on the network topology.

# On-going Experimental Project

A very recent thread of experimental papers tested non-Bayesian (DeGroot) vs. Bayesian models in the lab, e.g. Chandrasekhar et al. (2020), DeFilippis et al. (2022), and support boundedly rational models.

Molavi et al. (2017) provides axiomatization of DeGroot and Molavi (2022) offers test procedures to identify bounded rationality.

Building on these references, we are currently working on building and experiment to assess how our model performs against DeGroot's.

We build a **stochastic** model of **opinion exchange** in networks.

We show that we obtain **suitable properties** (convergence, consensus) **under weaker conditions** than standard models.

We further show that the limit belief is a full-support random variable: the model accounts for possible **large spread of marginal beliefs**.

We show that the expected **proportion of "wrong beliefs"** remains a constant **independent from the network structure**.

These two results strongly move away from usual non-Bayesian models.

On-going projects opened by this paper:

- Complete analytical characterization $\rightarrow$ exploiting probability coupling theory and Bayesian statistics results;
- Foundation of the belief-updating rule $\rightarrow$ what makes a good model of bounded rationality?;
- Extensions $\rightarrow$ metacognition and generalization of naïve learning;
- Building block in games of strategic belief manipulation $\rightarrow$ modelling large-scale informational phenomena;
- Experimental evaluation of the model $\rightarrow$ implementation in applied settings.

# Stochastic Consensus and the Shadow of Doubt

Emilien Macault, CREST & BETA Economics

Blockchain@X-OMI , Sept. 21st, 2023

# Notations

Denote by:

- $d_i$ agent $i$'s number of neighbors and $N(i)$ her neighborhood, $d = \min_i d_i$;
- $R_i^t$ and $B_i^t$ respectively the numbers of red and blue balls in player $i$'s urn at time $t$;
- $Z_i^t = R_i^t/(R_i^t + B_i^t)$ be the proportion of red balls in urn $i$ after step $t$;
- $X_i^t$ be the indicator variable of a red draw for agent $i$ at time $t$;
- $\mathcal{F}_t$ be the sigma-field generated by the realizations of $(X^k), k \leq t$

# Sketch of proof

We derive the following dynamics:

$$\begin{cases} R_i^{t+1} = R_i^t + \sum_{j\in N(i)} X_j^{t+1} \\ B_i^{t+1} = B_i^t + d_i - \sum_{j\in N(i)} X_j^{t+1} \\ R_i^{t+1} + B_i^{t+1} = 1 + d_i(t+1) \end{cases}$$

From which we get:

$$Z_i^{t+1} = Z_i^t + \frac{-d_i Z_i^t + \sum_{j\in N(i)} X_j^{t+1}}{1 + d_i(t+1)}$$

# Sketch of proof

Using $\mathbb{E}\left[X_i^{t+1}|\mathcal{F}_t\right] = Z_i^t$, we write:

$$Z_i^{t+1} - Z_i^t = \frac{\frac{1+d(t+1)}{1+d_i(t+1)}\left[-d_i Z_i^t + \sum_{j\in N(i)} Z_j^t + \sum_{j\in N(i)} X_j^{t+1} - \mathbb{E}\left[\sum_{j\in N(i)} X_j^{t+1}|\mathcal{F}_t\right]\right]}{1+d_i(t+1)}$$

$$\Leftrightarrow Z^{t+1} - Z^t = \gamma^t\left[f^t(Z^t) + u^t\right]$$

We identify a stochastic approximation algorithm.

# Sketch of proof

Observe the following:

1. We have $\begin{cases} \sum_{t=1}^{\infty} \gamma^t = \infty \\ \sum_{t=1}^{\infty} (\gamma^t)^2 < \infty \end{cases}$
2. For every $i$ in $N$, the sequence $(u_i^t)$ is a sequence of bounded random variables with zero mean $\rightarrow$ mart. diff. noise.
3. The maps $f_i^t$ are Lipschitz continuous and measurable with respect to $\mathcal{F}_t$ and uniformly continuous in $t$ for $t \geq 1$,
4. For any $z \in [0,1]^N$ and any $k \in \mathbb{N}^*$,

$$\lim_{s \to \infty} \left| \sum_{t=s}^{s+k} \gamma^t \left[ f_i^t(z) - \bar{f}_i(z) \right] \right| \rightarrow 0$$

with $\bar{f}_i(z) = \frac{d}{d_i} \sum_{j \in N(i)} z_j - d_i z_i$.

# Sketch of proof

**Theorem (Kushner & Yin (2003))**

*If observations 1-4 hold and $(Z^t)$ is bounded with probability one, then for almost all $\omega$, the limits $\bar{Z}(\omega)$ of convergent subsequences of $(Z^t(\omega))$ are trajectories of*

$$\dot{z}_i^t = \bar{f}(z^t) \tag{1}$$

*in some bounded invariant set and $(Z^t(\omega))$ converges to this invariant set.*

$\rightarrow$ $Z^t$ converges almost-surely for any initial vector $Z^0$ and any graph $G$.

# Sketch of proof

We know that $Z^t$ converges along trajectories of

$$\dot{z}_i^t = \frac{d}{d_i} \sum_{j \in N(i)} z_j^t - d_i z_i^t$$

As $\frac{1}{N} \leq \frac{d}{d_i} \leq 1$, stable points belong to the set of stable points of

$$\dot{z}_i^t = \sum_{j \in N(i)} z_j^t - d_i z_i^t$$

Or, in matrix form:

$$\dot{z}^t = -Lz^t$$

# Sketch of Proof

We have that $L = D - A$ with $D$ the diagonal matrix of degrees and $A$ the adjacency matrix of $G$.

In graph theory, it is refered to as the Laplacian matrix of $G$. It is always positive and semi-definite.

Additionally, the dimension of its nullspace if equal to the number of connected components of $G \rightarrow$ if $G$ is connected, this dimension is 1.

As the sum of line entries are all equal to zero, it is characterized by the eigenvector $(1, \ldots, 1)$.

# Sketch of Proof

We conclude that trajectories of the ODE converge to the nullspace of $-L$, which is Lyapunov stable as $-L$ is negative and characterized by the eigenvector $(1, \ldots, 1)$.

In other terms, for any initial condition, the sequence of proportions converges to a stable consensus.

Note that if $G$ is not connected, then this result applies within every connected component of $G$ (even for a isolated agent).

# Sketch of proof

The proof rests on the concept of *attainability* from Benaïm (1999).

**Definition**

A subset $I$ is attainable if for every fixed $T \geq 0$ there exists $t \geq T$ such that

$$\mathbb{P}\left(Z^t \in I\right) > 0.$$

Attainability rules out limit points that would be out of reach for the process.

# Sketch of proof

We adapt the following theorem from Renlund 2010:

**Lemma**

*Let $p$ be a stable zero of $\bar{f}$. If every neighborhood of $p$ is attainable then $p \in Supp(\bar{Z})$.*

For every $\epsilon > 0$, any $z \in [0,1]$ and every urns composition $Z^t$ with balls of the two colors, there exists a finite sequence of draws with positive probability such that for any $i \in \{1, \ldots, N\}$, $Z_i$ is at distance at most $\epsilon$ from $z$ (because $Z^{t+1} - Z^t$ is of the order $1/t$).

# $k$-regular graphs

**Theorem**

*If the graph $G$ is $k$-regular (every node has exactly $k$ neighbors) and connected, then* $\mathbb{E}\left[\bar{Z}|Z^0\right] = \alpha$.

# $k$-regular graphs

Consider a $k$-regular graph and let $M^t = \sum_{i \in N} Z_i^t$,

$$M_t - \mathbb{E}\left[M_{t+1}|\mathcal{F}_t\right] = \mathbb{E}\left[\sum_{i=1}^{N} Z_i^{t+1} - Z_i^t|\mathcal{F}_t\right] = \sum_{i=1}^{N} \frac{-d_i Z_i^t + \sum_{j \in N(i)} Z_j^t}{d_i\left(t+1\right)+1}$$

$$M_t - \mathbb{E}\left[M_{t+1}|\mathcal{F}_t\right] = \sum_{i=1}^{N} Z_i^t \left[\sum_{j \in N(i)} \frac{1}{d_j(t+1)+1} - \frac{1}{d_i(t+1)+1}\right]$$

# $k$-regular graphs

If $G$ is $k$-regular, $d_i = k$ for every $i$, hence

$$M_t - \mathbb{E}\left[M_{t+1}|\mathcal{F}_t\right] = \sum_{i=1}^{N} Z_i^t \left[\sum_{j\in N(i)} \frac{1}{k(t+1)+1} - \frac{1}{k(t+1)+1}\right] = 0$$

Which proves that $(M^t)$ is a martingale. In particular, we have that $\mathbb{E}\left[M_t|Z^0\right] = N\alpha$ for every $t$ hence $\mathbb{E}\left[\bar{Z}|Z^0\right] = \alpha$.