Distributed Estimation and Common Knowledge

Jeff Calder
Math 872
Queen's University

April 9, 2009

1 Introduction

Consider a system of M autonomous agents labeled $\{1, 2, \ldots, M\}$ and a random vector X which the agents wish to estimate. Such problems are called distributed estimation problems and have numerous practical applications such as in sensor networks or group decision problems. In a distributed estimation problem, each agent's estimate of X is updated when he makes an observation of X or when he receives a message from another agent. Thus each agent's estimate begins as a coarse initial guess and is then refined over time. Much of the research in distributed problems involves determining how each agent's estimate evolves over time and how to determine when two or more agents will agree about X.

This report aims to summarize two important papers addressing these issues. The first is *Agreeing to Disagree*[1], published by Robert Aumann in the journal *The Annals of Statistics* in 1976, and the second is *Asymptotic Agreement in Distributed Estimation*[2], published by Vivek Borkar and Pravin Varaiya in the journal IEEE transactions on Automatic Control in 1982. The first Section is an introduction to the concept of common knowledge and an attempt to motivate the results in Aumann's work. Next we bridge Aumann's work with that of Borkar and Varaiya and provide an introduction and summary of their work with some partial sketches of proofs. Finally we conclude with some discussion of possible improvements and open issues.

An attempt is made throughout the report to introduce and explain any mathematical concepts which the average reader may not be familiar with. The interested reader can find more thorough explanations in the referenced books on probability theory, particularly [3] and [4].

2 Common Knowledge

Two people, 1 and 2, are said to have *common knowledge* of an event A if both know it, 1 knows that 2 knows it, 2 knows that 1 knows it, 1 knows that 2 knows that 1 knows it, and so on. For example, if 1 and 2 are both present to witness the event A occurring and they can each clearly see the other, then the event becomes common knowledge. This idea is an important one in distributed estimation problems, game theory and other areas of applied mathematics. The above definition, although intuitive, cannot be easily adopted in any mathematical models. This was rectified by Robert Aumann's 1976 paper *Agreeing to Disagree* [1], which will be summarized in this section.

Let $(\Omega, \mathcal{B}, p)$ be a probability space. $(\Omega, \mathcal{B})$ represents the set of all possible states of the world and p the prior probability measure. This interpretation means that if an agent has no information about the state of the world, then for any event $A \in \mathcal{B}$ he concludes that the probability that A has occurred is $p(A)$. Before proceeding, we need a few definitions.

Definition 1. *A* ***decomposition*** *(or partition) of Ω is a disjoint covering of Ω by non-empty p-measurable sets.*

We should note that one can induce a partial ordering on the class of all decompositions. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two partitions of Ω. We say that $\mathcal{P}_1 \prec \mathcal{P}_2$ if for all $A \in \mathcal{P}_2$ there exists $B \in \mathcal{P}_1$ such that $A \subseteq B$. In other words, every element of $\mathcal{P}_2$ is contained in an element of $\mathcal{P}_1$, so $\mathcal{P}_2$ is a refinement of $\mathcal{P}_1$. Clearly, any two decompositions are not necessarily comparable, but we can always define a least upper bound (or max) and a greatest lower bound (or min). This allows us to define the following:

Lemma 1. *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two decompositions of Ω. Then:*

(i) *The least upper bound (or join) exists and is given by*

$$\mathcal{P}_1 \vee \mathcal{P}_2 = \left\{ A_1 \cap A_2 \neq \varnothing \;\middle|\; A_i \in \mathcal{P}_i, \quad i = 1, 2 \right\}$$

(ii) *The greatest lower bound (or meet) exists and is denoted $\mathcal{P}_1 \wedge \mathcal{P}_2$*

Remark 1. *We will also use the symbols $\vee$ and $\wedge$ to denote the similar operations on σ-fields. Given two σ-fields $\mathcal{F}$ and $\mathcal{G}$, $\mathcal{F} \vee \mathcal{G} = \sigma(\mathcal{F} \cup \mathcal{G})$ and $\mathcal{F} \wedge \mathcal{G} = \mathcal{F} \cap \mathcal{G}$.*

The notion of a decomposition will be essential in modeling how individual agents filter their information.

Definition 2. *An* ***information partition*** *is a decomposition $\mathcal{P}$ of Ω. If the state of the world is $\omega \in \Omega$, then the agent is informed of the partition element containing ω denoted $\mathcal{P}(\omega)$.*

For a given state $\omega \in \Omega$, we can consider what it means for an agent to have knowledge of some event $A \in \mathcal{B}$. If an agent has knowledge of $A \in \mathcal{B}$ then the agent should know with certainty (i.e. with probability one) that the event occurred. Thus, the conditional probability $p(A|\mathcal{P}(\omega))$ should be equal to one. In other words:

$$p(A|\mathcal{P}(\omega)) = \frac{p(A \cap \mathcal{P}(\omega))}{p(\mathcal{P}(\omega))} = 1$$

which is only true when $p(A \cap \mathcal{P}(\omega)) = p(\mathcal{P}(\omega))$ which means $\mathcal{P}(\omega) \subset A$ (or $\mathcal{P}(\omega) - N \subset A$ for a measure zero set $N \subset \mathcal{P}(\omega)$, but we will ignore this possibility for now). This motivates the following definition.

Definition 3. *We say an agent has **knowledge of A at** ω if $\mathcal{P}(\omega) \subset A$.*

We also sometimes say that the agent *knows* A. Note that this definition depends on the state $\omega \in \Omega$. Now consider two agents, 1 and 2, with information partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively and assume that $A \in \mathcal{B}$ is common knowledge for 1 and 2 at some $\omega \in \Omega$. It is an implicit assumption that both 1 and 2 know how information is imparted to each other, thus 1 knows 2's information partition $\mathcal{P}_2$ and vice versa. Since A is common knowledge at ω, both 1 and 2 know it, thus $\mathcal{P}_1(\omega) \subset A$ and $\mathcal{P}_2(\omega) \subset A$. Now consider the statement *1 knows that 2 knows* A. Now, 1 does not know precisely which partition element, $\mathcal{P}_2(\omega) \in \mathcal{P}_2$, 2 was informed of so we must consider all the possibilities. Since the state must be some $\omega \in \mathcal{P}_1(\omega)$, 1 knows that $\mathcal{P}_2(\omega)$ must intersect $\mathcal{P}_1(\omega)$. So if 1 knows that 2 knows A, then A must contain any $P \in \mathcal{P}_2$ which intersects $\mathcal{P}_1(\omega)$ or else 1 could never be certain whether or not 2 knows A. The argument is similar for agent 1. Thus A must contain all partition elements of $\mathcal{P}_2$ which intersect $\mathcal{P}_1(\omega)$ and all partition elements of $\mathcal{P}_1$ which intersect $\mathcal{P}_2(\omega)$. Continuing this logic for the statement *1 knows that 2 knows that 1 knows* A and so on, one can see that A must contain all sequences of partition elements $P^1, P^2, \ldots, P^n$ where $P^1 = \mathcal{P}_1(\omega)$ or $P^1 = \mathcal{P}_2(\omega)$, the partition elements P^i belong alternatively to $\mathcal{P}_1$ and $\mathcal{P}_2$ (i.e. $P^i \in \mathcal{P}_1$, $P^{i+1} \in \mathcal{P}_2$, $P^{i+2} \in \mathcal{P}_1 ...$) and $P^i \cap P^{i+1} \neq \varnothing$. This implies (with a bit of thought) that A must contain the element of the greatest lower bound $\mathcal{P}_1 \wedge \mathcal{P}_2$ which contains ω. The main contribution from [1] is the theorem/definition:

Theorem 1. *Given that the state of the world is $\omega \in \Omega$, an event A is said to be **common knowledge at** ω for agents 1 and 2 if A contains the partition element of $\mathcal{P}_1 \wedge \mathcal{P}_2$ corresponding to ω.*

Given this definition, Aumann goes on to show that if two people have the same priors for an event A and their posteriors for A are common knowledge then they must be equal, hence they cannot agree to disagree. The proof of this theorem is immediate from the above theorem. The importance of this work is the casting of common knowledge into a useful set theoretic framework. In the next section we will discuss generalizations of this idea.

2.1 Extensions of Aumann

The discussion in the previous section assumed a given state of the world $\omega \in \Omega$ and asked what it means for an event to be common knowledge for two agents at ω. Now, if we try to ask a slightly different question, such as *what it means for an event $A \in \mathcal{B}$ to be common knowledge for agents 1 and 2 for every state $\omega \in \Omega$*, we get a slightly unsatisfactory answer. Using theorem 1, this means that every partition element $P \in \mathcal{P}_1 \wedge \mathcal{P}_2$ is contained in A thus $A = \Omega$.

It turns out that the appropriate question to ask is: *for an event A, what does it mean for the occurrence or non-occurrence of A to be common knowledge for every $\omega \in \Omega$?* Thus for every $\omega \in \Omega$, either A or A^c is common knowledge for two agents 1 and 2. Using theorem 1 this implies that for every $P \in \mathcal{P}_1 \wedge \mathcal{P}_2$, either $P \subset A$ or $P \subset A^c$. It is easy to see that this forces A to be a union of partition elements of $\mathcal{P}_1 \wedge \mathcal{P}_2$. As an aside, note that given any partition $\mathcal{P}$ of Ω one can define a σ-field $\sigma(\mathcal{P})$ by taking all countable unions of partition elements. Furthermore, given two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ their respective σ-fields $\sigma(\mathcal{P}_1)$ and $\sigma(\mathcal{P}_2)$, it is not hard to see that

$$\sigma\left(\mathcal{P}_1 \wedge \mathcal{P}_2\right) = \sigma(\mathcal{P}_1) \cap \sigma(\mathcal{P}_2)$$

This leads us the first generalization of Aumann's common knowledge.

Definition 4. *An event $A \in \mathcal{B}$ is **common knowledge** for 1 and 2 if $A \in \sigma(\mathcal{P}_1 \wedge \mathcal{P}_2)$ (equivalently 1_A is measurable on $\sigma(\mathcal{P}_1)$ and $\sigma(\mathcal{P}_2)$).*

Thus measurability seems to be the correct tool to model common knowledge. This observation allows us to extend the idea of common knowledge to random variables.

From this point on, we will consider σ-fields as information structures in place of the more primitive notion of information partitions. Since every information partition generates a σ-field (but not vice versa) there is no loss of generality. Thus if we have a group of M agents $\{1, 2, \ldots, M\}$ then we will associate σ-fields $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_M$ (for each i, $\mathcal{F}_i \subset \mathcal{B}$) with the information available to each agent. Then we have the following generalization of common knowledge.

Definition 5. *A random variable $X : \Omega \to \mathbb{R}$ is said to be **common knowledge** for agents $\{1, 2, \ldots, M\}$ if X is measurable with respect to each σ-field $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_M$.*

3 Distributed Estimation

Suppose we have a set $A := \{1, 2, 3, \dots, M\}$ of M agents and a random vector X which the agents wish to estimate. Furthermore, suppose each agent receives signals at random times from the environment and that these signals contain some information about X. After agent m receives a signal, he can update his estimate of the random variable X with whatever new information was contained in the signal. In such a scenario, there is obviously no reason for any of the agents to agree on their estimates of X (although they may do so). Now suppose the agents can communicate with each other. Suppose that at random times m transmits her estimate of the state vector to a random subset of other agents. Using this information, the other agents can then update their estimates accordingly and then transmit these to other agents and the process continues until an equilibrium is reached. Again, it is not necessary that all the agents agree; indeed, some may never even communicate. But we can ask the question: under what conditions will all (or a subset) of the agents agree on the estimate of X? This is the question answered by [2] in the paper *Asymptotic Agreement in Distributed Estimation* and there is a very interesting connection between this answer and the concept of common knowledge discussed in the previous sections. In this section we will attempt to set up the problem and understand the results in [2].

Let $\mathbb{R}_+$ denote the set of non-negative real numbers and let X be an integrable random vector.

Definition 6. *Let $(\Omega, \mathcal{F}, p)$ be a probability space, $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a filtration (i.e. a family of σ-fields such that $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ for all $s < t$) and $(X_t)_{t\in\mathbb{R}_+}$ be a family of random variables adapted to the family $(\mathcal{F}_t)$ (i.e. X_t is measurable with respect to $\mathcal{F}_t$ for all $t \in \mathbb{R}$). Then (X_t) is a **continuous time martingale** with respect to the family $(\mathcal{F}_t)$ if:*

(i) $E|X_t| < \infty \;\; \forall t \in \mathbb{R}_+$

(ii) $E[X_t|\mathcal{F}_s] = X_s \; \forall s, t \in \mathbb{R}_+$ *such that* $s \le t$

Many of the discrete-time martingale theorems carry over to continuous-time provided $(\mathcal{F}_t)$ is a complete right continuous family. (F_t) is right continuous if

$$\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$$

where the right hand side of the above equation can be thought of as the limit from the right. $(\mathcal{F}_t)$ is a complete family if each $\mathcal{F}_t$ contains all sets of probability zero. Given a continuous time martingale, one can (almost) always complete the filtration and obtain a complete right continuous family. For more details the reader is referred to [3]. We will assume from now on that every filtration is complete and right continuous.

Let $\mathcal{F}_t^m$ be the filtration representing the information available to m at time t. Let X_t^m be m's best estimate of the random vector X at time t. If we assume that each agent m is trying to minimize a cost function of the form

$$J(X_t^m) = E\left[(X - X_t^m)^2 \,|\mathcal{F}_t^m\right]$$

Then one can show that X_t^m is given by

$$X_t^m = E[X|\mathcal{F}_t^m]$$

The above equation implies that $(X_t^m, \mathcal{F}_t^m)$ is a martingale. To see this, note that by definition of conditional expectation, X_t^m is measurable on $\mathcal{F}_t^m$ for every $t \in \mathbb{R}_+$ and using the law of iterated expectations, we have

$$E\left[|X_t^m|\right] \leq E\left[E[\,|X|\,|\mathcal{F}_t^m]\right] = E|X| < \infty$$

where we used the fact that $|X_t^m| \leq E[|X| \mid \mathcal{F}_t^m]$ almost surely. Finally, for $s < t$ we have $\mathcal{F}_s^m \subset \mathcal{F}_t^m$, so using the law of iterated expectations again we have

$$E[X_t^m|\mathcal{F}_s^m] = E[E[X|\mathcal{F}_t^m]|\mathcal{F}_s^m] = E[X|\mathcal{F}_s^m] = X_s^m$$

Thus m's estimate of X is a martingale.

Theorem 2. $(X_t^m, \mathcal{F}_t^m)$ *is a martingale.*

The most natural question to ask is: Does each each agent's estimate X_t^m converge to something as $t \to \infty$? From an above calculation, we have

$$\sup_{t \in \mathbb{R}_+} E[|X_t^m|] \leq E|X| < \infty$$

Thus the martingale convergence theorem guarantees that each agent's estimate X_t^m converges almost surely to some random variable as $t \to \infty$. Let us define

$$X_\infty^m := \lim_{t \to \infty} X_t^m \quad \text{a.s.}$$

Furthermore, one can show that the family (X_t^m) is uniformly integrable. Using the fact that $|X_t^m| \leq E[|X| \mid \mathcal{F}_t^m]$ almost surely we have

$$\begin{aligned} E\left[|X_t^m|\, 1_{\{|X_t^m|>c\}}\right] &\leq E\left[E\left[|X| \mid \mathcal{F}_t^m\right] 1_{\{|X_t^m|>c\}}\right] \\ &= E\left[E\left[|X| 1_{\{|X_t^m|>c\}} \mid \mathcal{F}_t^m\right]\right] \\ &= E\left[|X| 1_{\{|X_t^m|>c\}}\right] \\ &= \int 1_{\{|X_t^m(\omega)|>c\}} |X(\omega)| dP(\omega) \end{aligned}$$

And by Markov's inequality we have that

$$P(\{|X_t^m(\omega)| > c\}) \leq \frac{E[|X_t^m|]}{c} \leq \frac{E[|X|]}{c}$$

Combining the above with the fact that X is an integrable random vector, one can show (with some work) that $E\left[|X_t^m|\, 1_{\{|X_t^m|>c\}}\right] \to 0$ as $c \to \infty$ uniformly in t. This implies that the family (X_t^m) is uniformly integrable and therefore X_∞^m closes the martingale on the right, i.e.

$$X_t^m = E\left[X \mid \mathcal{F}_t^m\right] = E\left[X_\infty^m \mid \mathcal{F}_t^m\right]$$

Now that we know each agent's estimate converges, a more interesting question is: For two agents m and p, are there certain conditions under which $X_\infty^m = X_\infty^p$? To answer this question, one needs to explicitly consider the messages transmitted between agents (note that this is implicitly contained in $\{\mathcal{F}_t^m\}$). This requires a further extension of our model.

Let t_j^m be an increasing sequence of $(\mathcal{F}_t^m)$-stopping times. At each t_j^m, m transmits his estimate $X_{t_j^m}^m$ to a random subset of agents $A_j^m \subset A$. Suppose that m receives signals from the environment and from other agents at random times. Let r_j^m be the $(\mathcal{F}_t^m)$-stopping times at which m receives a message from another agent and call this message Z_j^m. We will assume that all transmission times are finite ($t_j^m < \infty$) and that every agent sends messages infinitely often ($\lim_{j\to\infty} t_j^m = \infty$).

Define a *transception* at m to be either a transmission or a reception at m. Let τ_n^m be the transception times for agent m. Since the reception and transmission times are stopping times, so are the transception times. Let $\mathcal{G}_n^m$ be the σ-field generated by the first n transceptions at m. Thus $\mathcal{G}_n^m$ contains the information from the estimates transmitted by m and received by m up to the n^{th} transception, but does not explicitly include any information contained in the signals from the environment.

Let $r_k^m(p)$ be the time at which agent m receives the k^{th} message from agent p and let $Z_k^m(p)$ be this message. Also, assume that each agent learns the identity of the transmitting agent whenever he receives a message. Set $r_k^m(p) = \infty$ if m receives fewer than k messages from p. Similarly, let $t_k^m(p)$ be the times at which m sends messages to p and let $X_k^m(p)$ be these messages. Again, set $t_k^m(p) = \infty$ if m sends fewer than k messages to p. Finally, let $q_k^m = n$ if the k^{th} transmission from m to any agent occurs at the n^{th} transception time. q_n^m is clearly a $(\mathcal{G}_n^m)$-stopping time. Lastly, assume that transmission delays are finite and no transmissions are lost or corrupted. That is:

$$\{\omega \mid t_k^m(p) < \infty\} = \{\omega \mid r_k^p(m) < \infty\} \tag{3.1}$$

for all k.

Note that since $\mathcal{G}_{q_k^m}^m$ represents the information generated by transceptions at m, it includes the information transmitted by m at the stopping time q_k^m which is $X_{q_k^m}^m$. Thus $X_{q_k^m}^m$ is measurable on $\mathcal{G}_{q_k^m}^m \subset \mathcal{F}_{q_k^m}^m$ so we have

$$X_{q_k^m}^m = E[X|\mathcal{F}_{q_k^m}^m] = E[X|\mathcal{G}_{q_k^m}^m]$$

Since (X_t^m) is a uniformly integrable family, the optional sampling theorem guarantees that $(X_{t_j^m}^m, \mathcal{F}_{t_j^m}^m)$ and $(X_{q_k^m}^m, \mathcal{G}_{q_k^m}^m)$ are martingales and

$$\lim_{k\to\infty} X_{q_k^m}^m = X_\infty^m \quad \text{and} \quad \lim_{j\to\infty} X_{t_j^m}^m = X_\infty^m$$

Let $\mathcal{G}_\infty^m$ be the limit σ-algebra of the filtration $\mathcal{G}_{q_k^m}^m$. That is

$$\mathcal{G}_\infty^m = \bigvee_{k\in\mathbb{N}} \mathcal{G}_{q_k^m}^m$$

From the above convergence results, we have that X_∞^m is measurable on $\mathcal{G}_\infty^m$.

Now that the model is fully described, we can obtain some immediate results. Let S^{mp} be the event that m sends his estimate to p infinitely often. We can write S^{mp} as

$$\begin{aligned} S^{mp} &= \bigcap_{k=1}^{\infty} \{\omega | t_k^m(p) < \infty\} \\ &= \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \{\omega |\ p \in A_j^m \text{ and } t_j^m < \tau_n^m \text{ for at least k distinct j}\} \end{aligned}$$

Since for any $k, n \in \mathbb{N}$, the set in the last line is measurable on $\mathcal{G}_n^m$ we have that $\{\omega | t_k^m(p) < \infty\}$ and hence S^{mp} are measurable on $\mathcal{G}_\infty^m$. Similarly, one can express S^{mp} as:

$$
\begin{aligned}
S^{mp} &= \bigcap_{k=1}^{\infty} \{\omega | r_k^p(m) < \infty\} \\
&= \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \{\omega | \ r_j^p < \tau_n^p \ \text{ for at least k distinct messages from m}\}
\end{aligned}
$$

and obtain that $\{\omega | r_k^p(m) < \infty\}$ and S^{mp} are measurable on $\mathcal{G}_\infty^p$. Thus $1_{\{S^{mp}\}}$ is common knowledge for $\mathcal{G}_\infty^m$ and $\mathcal{G}_\infty^p$. This agrees with our intuition, because both agents know who they are transmitting to and who they are receiving messages from. Thus, the event that they communicate infinitely often should be common knowledge.

Now, the transmission delays are assumed finite but we have not assumed that messages are received in the same order as they are transmitted. Suppose that the k^{th} message transmitted from m to p is actually the ℓ_k^{th} message received at p from m. Then the message received by p, $Z_{\ell_k}^p(m)$ is exactly $X_{t_k^m(p)}^m$. Using this and equation (3.1) we have the following equality

$$
X_{t_k^m(p)}^m 1_{\{\omega | t_k^m(p) < \infty\}} = Z_{\ell_k}^p(m) 1_{\{\omega | r_k^p(m) < \infty\}}
$$

This is a key equality as it provides a link between X_∞^m and X_∞^p. $\{\omega | t_k^m(p) < \infty\}$ monotonically decreases to the set S^{mp} as $k \to \infty$ and thus

$$
\lim_{k \to \infty} X_{t_k^m(p)}^m 1_{\{\omega | t_k^m(p) < \infty\}} = X_\infty^m 1_{\{S^{mp}\}} \quad \text{a.s.}
$$

Similarly, $1_{\{\omega | r_k^p(m) < \infty\}}$ monotonically decreases to S^{mp} and on this set, m sends messages infinitely often to p, so $\ell_k \to \infty$ as $k \to \infty$, so we have

$$
\begin{aligned}
\lim_{k \to \infty} Z_{\ell_k}^p(m) 1_{\{\omega | r_k^p(m) < \infty\}} &= \lim_{k \to \infty} X_{t_k^m(p)}^m 1_{\{\omega | t_k^m(p) < \infty\}} \\
&= X_\infty^m 1_{\{S^{mp}\}} \quad \text{a.s.}
\end{aligned}
$$

Since each term in the sequence $Z_{\ell_k}^p(m) 1_{\{\omega | r_k^p(m) < \infty\}}$ is measurable on $\mathcal{G}_k^p$, we obtain that $X_\infty^m 1_{\{S^{mp}\}}$ is measurable on $\mathcal{G}_\infty^p$. This result is summarized as the following theorem.

Theorem 3 (Borkar)**.** *$X_\infty^m 1_{\{S^{mp}\}}$ and $1_{\{S^{mp}\}}$ are common knowledge for $\mathcal{G}_\infty^m$ and $\mathcal{G}_\infty^p$.*

This theorem has some very interesting consequences. Note that the theorem implies that $X_\infty^m 1_{\{S^{mp}\}}$ and $1_{\{S^{mp}\}}$ are measurable on the σ-field $\mathcal{G}_\infty^m \cap \mathcal{G}_\infty^p$. By symmetry, $X_\infty^p 1_{\{S^{pm}\}}$ and $1_{\{S^{pm}\}}$ are also measurable on $\mathcal{G}_\infty^m \cap \mathcal{G}_\infty^p$. Let $S = S^{mp} \cap S^{pm}$, thus S denotes the event that m and p communicate infinitely often. It follows that $X_\infty^m 1_{\{S\}}$ is measurable on $\mathcal{G}_\infty^m \cap \mathcal{G}_\infty^p$. Thus we can write

$$\begin{aligned} X_\infty^m 1_{\{S\}} &= E[X|\mathcal{G}_\infty^m]1_{\{S\}} \\ &= E[X1_{\{S\}}|\mathcal{G}_\infty^m] \\ &= E[X1_{\{S\}}|\mathcal{G}_\infty^m \cap \mathcal{G}_\infty^p] \end{aligned}$$

The above expression for $X_\infty^m 1_{\{S\}}$ is symmetric in m and p so it is equally true for $X_\infty^p 1_{\{S\}}$ thus we obtain the following corollary.

Corollary 1 (Borkar). $X_\infty^m 1_{\{S\}} = X_\infty^p 1_{\{S\}}$ *a.s.*

The above corollary says that on any sample path ω where m and p communicate infinitely often, $X_\infty^m(\omega)$ and $X_\infty^p(\omega)$ must agree (i.e. m and p agree asymptotically).

At this point, it is natural to ask if this result holds for two agents who communicate infinitely often through other agents rather than directly. In the case of two agents, both agents knew they were communicating infinitely often with each other, and more importantly each agent knew that the other knew they were communicating infinitely often. If we consider the case of three or more agents, it is possible that 2 agents may communicate infinitely often through a third agent without knowing that they do so. Furthermore, even if each agent were aware that they were communicating indirectly, it is still possible that neither agent will know that the other knows about their communication. In the case of multiple agents, it turns out that the concept of common knowledge it the key to understanding when we will have asymptotic agreement.

Definition 7. *A sequence of agents $m_1, m_2, \ldots, m_n, m_{n+1} = m_1$ forms a communicating ring for an event $S \in \mathcal{F}$ if $S \subset S^{m_i m_{i+1}}$ for $i = 1, 2, 3, \ldots, n$.*

Thus each agent in a communicating ring communicates infinitely often with every other member either directly or indirectly through another agent. The following lemma will be prove useful in proving a result about asymptotic agreement for communicating rings. For a proof of this lemma, see [2].

Lemma 2. *Let $W_1, \dots, W_n, W_{n+1} = W_1$ be integrable random vectors and $H_1, \dots, H_n$ be σ-fields such that $W_i = E[W_{i+1}|H_i]$, $i = 1, \dots, n$. Then $W_1 = \cdots = W_n$ a.s.*

The following theorem extends the results about asymptotic agreement from two agents to a communicating ring of n agents.

Theorem 4 (Borkar)**.** *Suppose that*

(i) *$m_1, m_2, \dots, m_{n+1} = m_1$ forms a communicating ring for $1_{\{S\}}$*

(ii) *$1_{\{S\}}$ is common knowledge for $\mathcal{G}_\infty^{m_1}$, $\mathcal{G}_\infty^{m_2}$, $\dots$, $\mathcal{G}_\infty^{m_n}$*

Then $X^{m_1} 1_{\{S\}} = X^{m_2} 1_{\{S\}} = \cdots = X^{m_n} 1_{\{S\}}$ a.s.

Proof. We will present the main steps in the proof. From theorem 3 and the fact that $\mathcal{G}_\infty^{m_i} \cap \mathcal{G}_\infty^{m_{i+1}} \subset \mathcal{G}_\infty^{m_i}$ we can write

$$\begin{aligned} X_\infty^{m_i} 1_{\{S^{m_i, m_{i+1}}\}} &= E[X|\mathcal{G}_\infty^{m_i} \cap \mathcal{G}_\infty^{m_{i+1}}] 1_{\{S^{m_i, m_{i+1}}\}} \\ &= E[E[X|\mathcal{G}_\infty^{m_{i+1}}]|\mathcal{G}_\infty^{m_i} \cap \mathcal{G}_\infty^{m_{i+1}}] 1_{\{S^{m_i, m_{i+1}}\}} \\ &= E[X_\infty^{m_{i+1}}|\mathcal{G}_\infty^{m_i} \cap \mathcal{G}_\infty^{m_{i+1}}] 1_{\{S^{m_i, m_{i+1}}\}} \\ &= E[X_\infty^{m_{i+1}} 1_{\{S^{m_i, m_{i+1}}\}}|\mathcal{G}_\infty^{m_i} \cap \mathcal{G}_\infty^{m_{i+1}}] \end{aligned}$$

Since by hypothesis (i), $S \subset S^{m_i, m_{i+1}}$ and by hypothesis (ii), $S \in \mathcal{G}_\infty^{m_i}$ for all i, we can multiply both sides by $1_{\{S\}}$ to get

$$X_\infty^{m_i} 1_{\{S\}} = E[X_\infty^{m_{i+1}} 1_{\{S\}}|\mathcal{G}^{m_i} \cap \mathcal{G}^{m_{i+1}}]$$

Letting $W_i = X_\infty^{m_i} 1_{\{S\}}$ and $H_i = \mathcal{G}^{m_i} \cap \mathcal{G}^{m_{i+1}}$, the result follows directly from lemma 2. □

Using the properties of conditional expectation, it is not difficult to show that because $X_\infty^{m_1} 1_{\{S\}} = X_\infty^{m_2} 1_{\{S\}} = \cdots = X_\infty^{m_n} 1_{\{S\}}$, each agent's estimate is given by

$$X_\infty^{m_i} 1_{\{S\}} = E[X|\mathcal{G}_\infty^{m_1} \cap \cdots \cap \mathcal{G}_\infty^{m_n}] 1_{\{S\}}$$

An important corollary is

Corollary 2 (Borkar)**.** *Suppose that with probability one all the agents form a communicating ring. Then the estimate of each agent converges to $E[X|\mathcal{G}_\infty^1 \cap \cdots \cap \mathcal{G}_\infty^M]$ almost surely.*

The first hypothesis in theorem 4 states that every pair of agents in the communicating ring sends messages to each other infinitely often either directly or indirectly through other agents in the ring. The second hypothesis states that for an event S, all of the agents know that they are members of the communicating ring. It is important to note that the first hypothesis does not imply the second. We could have agents m_1, m_2, m_3 where m_1 sends messages to m_2 infinitely often, m_2 to m_3 i.o., and then m_3 to m_1 i.o.. Although they form a communicating ring, m_1 may not be able to infer whether m_2 sends messages to m_3 i.o. In fact one can find counter examples where members of a communicating do not agree asymptotically when the second hypothesis is violated. The reader is referred to [2] for more details.

3.1 (Non)Optimality of the Asymptotic Estimate

Assume that with probability one, all agents form a communicating ring. Then all agents agree asymptotically and their estimates converge to $\hat{X}$ given by

$$\hat{X} = E[X|\mathcal{G}^1_\infty \cap \mathcal{G}^2_\infty \cap \cdots \cap \mathcal{G}^M_\infty]$$

Now suppose the agents were to share their raw information rather than their processed estimates of X. Such a situation could arise if all the agents were to send their information to some central processor which would then make the estimate of X. Define $\mathcal{F}^m_\infty$ by

$$\mathcal{F}^m_\infty = \bigvee_{t \geq 0} \mathcal{F}^m_t$$

Then the full information estimate X^* (i.e. given all the raw information from all agents) is

$$X^* = E[X|\mathcal{F}^1_\infty \vee \mathcal{F}^2_\infty \vee \cdots \vee \mathcal{F}^M_\infty]$$

Since more information is considered when computing X^*, it is clear that X^* will be at least as good an estimate as $\hat{X}$. In fact, in some situations, X^* can be strictly better than $\hat{X}$. The following example from [2] illustrates this.

Example 1. *Take* $\Omega = [0,1] \times [0,1]$ *and* p *to be Lebesgue measure. Let* $A \subset \Omega$ *be the hashed region in figure 1 and let* $X = 1_{\{A\}}$ *be the random variable agents 1 and 2 wish to estimate. Suppose that agent 1 observes* $1_{\{B\}}$ *and agent 2 observes* $1_{\{C\}}$*. Then each agent will independently form the estimate* $X^1_t = X^2_t = E[X] = 0.5 \;\text{ a.s.}$ *Furthermore, if the agents transmit*

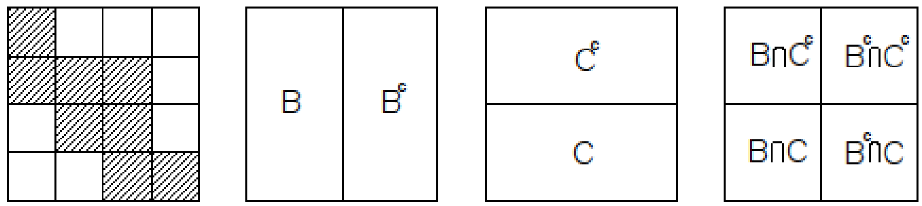

Figure 1: Figures for example 1

their estimates to each other, neither will change his or her estimate since they already agree. Now suppose the agents share their raw information instead of just their estimates of X. *Then the agents will be able to observe* $1_{\{B \cap C\}}$,$1_{\{B \cap C^c\}}$,$1_{\{B^c \cap C\}}$ *and* $1_{\{B^c \cap C^c\}}$. *Thus the full information estimate* X^* *will be the random variable that takes values* 0.75 *and* 0.25 *each with probability* 0.5. *Thus* X^* *can be strictly better than* $\hat{X}$.

4 Discussion and Concluding Remarks

The paper *Agreeing to Disagree* [1] is a very popular and seminal paper which set the initial framework for many other works in distributed estimation problems, including Borkar and Varaiya [2]. Aumann's approach is simplistic but his important contribution is the set theoretic framework for the concept of common knowledge, which had not been previously studied in a mathematical framework. Generalizations of his ideas to the more powerful language of measure theory are, as we have seen, immediate. The paper by Borkar and Varaiya [2] can be seen as a significant extension of Aumann's "agreeing to disagree" result. For the case of two agents, the asymptotic agreement result is very similar to Aumann's "agreeing to disagree" result except cast in the language of measure theory. The case for multiple agents could be made simple as well, if we merely assume that all agents communicate with all other agents infinitely often. However this is a bit simplistic and unrealistic. The result in [2] is much stronger than this. All that is required is that a group of agents forms a communicating ring, and that the event that they form the ring is common knowledge for all the ring members.

There are a few unanswered questions posed by the authors of [2]. Firstly, although we have conditions under which all (or a subset) of agents will agree asymptotically, in any practical application decisions have to be made in finite time. It would thus be useful to get bounds on the rate of convergence. Secondly, since the full information estimate X^* is always better than $\hat{X}$, it would seem that having the agents share their raw data is better than sharing their estimates (processed data).

The first point is certainly important for any applications. One would suspect that the rate of convergence would depend on the size of the communicating ring and how often the members of the ring exchange estimates of X. However, since neither of these factors are incorporated into the model as described in [2] any satisfactory answer to the first question would require additional constraints. As far as I know, this is an open problem in the general case.

The second point raises another interesting question about the role of a central processor in a distributed estimation problem. In order to make the most accurate estimate of X, all agents should be sharing their raw data with each other or preferably with a central processor which would then make the full information estimate. However, this is at ends with practical considerations as sharing raw data would likely use more resources (power, bandwidth, etc) than first pre-processing the data and then sharing it in a more compact form. Perhaps a solution is for agents to share their estimates, but also their raw data whenever possible depending on some measure of their available resources (e.g. battery life...). It is not clear how to analyze such a system, but one would expect a better estimate because at least some raw data will be exchanged.

In conclusion, the papers *Agreeing to Disagree* and *Asymptotic Agreement in Distributed Estimation* provide some very interesting answers to problems concerning distributed estimation and have found applications in numerous fields such as game theory and economics. Distributed estimation problems are abundant in the world around us and distributed problems are still an open field of research today.

References

[1] Robert J. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976.

[2] V. Borkar and P. Varaiya. Asymptotic agreement in distributed estimation. 27(3):650–655, Jun 1982.

[3] Claude Dellacherie and Paul-Andre Meyer. *Probabilities and Potential.* North-Holland Publishing Company, 1982.

[4] Kiyosi Itô. *Introduction to Probability Theorey.* Cambridge University Press, 1986.

[5] Serdar Yüksel. *Math 872 Course Lecture Notes.* Queen's University, 2009.