ASYMPTOTIC AGREEMENT IN DISTRIBUTED ESTIMATION

by

V. Borkar and P. Varaiya

Memorandum No. UCB/ERL M80/45

23 September 1980

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Asymptotic Agreement in Distributed Estimation

Vivek Borkar[1] and Pravin Varaiya

Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California, Berkeley, California 94720

Abstract

Each of several agents updates his estimate of the same random variable whenever he makes a new observation or receives the estimate made by another agent. In turn, each agent transmits his estimate to a randomly chosen subset of the other agents. A subset of agents forms a communicating ring if for every pair m, p of ring members there is a sequence of ring members $m = m_1, m_2, \ldots, m_{n+1} = p$ such that m_i sends his estimate to m_{i+1} infinitely often. If each ring member knows that he is a ring member, then the estimates of all the ring members asymptotically agree. However this common limit can depend upon the order in which estimates are transmitted.

Research supported in part by the Office of Naval Research under Contract N00014-80-C-0507.

[1]Current address: Dept. of Applied Mathematics, Twente University of Technology, Post box 217, 7500 AE Enschade, Netherlands.

1. Introduction

Consider a set of agents $A := \{1,..,M\}$ and a fixed random vector X which each agent wishes to estimate. If at any time t, the information available to $\underline{m}$ is represented by the σ-field $\underline{\underline{F}}_t^m$, then m's estimate of X is assumed to be $X_t^m := E\{X|\underline{\underline{F}}_t^m\}$. At random times $\underline{m}$ transmits his current estimate to a randomly selected subset of the other agents. In turn, his own information field $\underline{\underline{F}}_t^m$ is generated by two sequences of 'observations'. The first consists of 'signals' received from the environment. The second consists of 'messages' received from the other agents; these messages are themselves the estimates of the transmitting agents. Both sequences of observations are received at random times, and a message sent by any agent is subject to a random transmission delay. We wish to study the convergence of the estimates X_t^m as the number of messages exchanged grows without bound. Simpler versions of this process of message exchange and estimation were investigated by Aumann [1] and Geanakoplos and Polemarchakis [2]. The results derived below extend theirs.

It is easy to see that X_t^m converges to a random variable X_∞^m as $t \to \infty$. The more interesting question is to determine the sample paths ω along which two different agents $\underline{m}$ and $\underline{p}$ agree asymptotically i.e. $X_\infty^m(\omega) = X_\infty^p(\omega)$. The answer can be informally described as follows. Say that $\underline{m}$ and $\underline{p}$ communicate infinitely often (i.o) along ω if there is a sequence of agents $m_1,\ldots,m_n,m_{n+1} = m_1$ such that $m = m_i$, $p = m_j$ for some i and j, and m_k sends messages to m_{k+1} i.o. along ω. Then $X_\infty^m(\omega) = X_\infty^p(\omega)$ if $\underline{m}$ and $\underline{n}$ communicate i.o. and they know that they do so. (A counterexample is given to show that the first condition alone is insufficient for asymptotic agreement.) In particular, if all agents communicate i.o. with each other and with probability one, then there is universal asymptotic agreement. This common limit however does generally depend upon the order in which the agents transmit their estimates. Generally it also differs from the full information or centralized estimate of X which would be obtained if all agents shared all their "raw data" and not just their estimates.

Similar results hold for a distributed detection problem in which the agents have to decide which of two alternative probability distributions correctly accounts for the observation processes. In this problem the message transmitted by an agent is his current estimate of the likelihood ratio.

In the concluding section we argue that these results suggest that in the design of a distributed information system, different agents should share their raw data as much as possible instead of individually processing it first and then sharing the resulting estimates.

2. The model

$(\Omega, \underline{F}, P)$ is the underlying probability space and X is a fixed random vector with $E|X| < \infty$. Each agent $\underline{m}$, $1 \leq \underline{m} \leq M$ operates as follows.

(i) $\underline{m}$ receives signal Z_j^{m0} at random time r_j^{m0}, $j = 1,2,\ldots$, with $0 \leq r_j^{m0} \leq r_{j+1}^{m0} \leq \infty$ a.s.

(ii) $\underline{m}$ receives message Z_j^m at random time r_j^m, $j = 1,2,\ldots$, with $0 \leq r_j^m \leq r_{j+1}^m \leq \infty$ a.s. Here Z_j^m is the estimate transmitted to $\underline{m}$ by some other agent. It is assumed that when a message is received, $\underline{m}$ also learns the identity of the transmitting agent.

Let $\hat{\underline{\underline{F}}}_t^m$ denote the σ-field of the information available to $\underline{m}$ at time t. Thus $\hat{\underline{\underline{F}}}_t^m$ is generated by subsets of the form

$$\{Z_j^{m0} \in B\} \cap \{r_j^{m0} \leq s\}\ ,\ \{Z_j^m \in B\} \cap \{r_j^m \leq s\}\ ,$$

where B is a Borel set and $s \leq t$. Note that some information may be conveyed by the time instance at which signals and messages arrive. Let $(\underline{F}_t^m, t \geq 0)$ denote a complete, right-continuous version of the family $(\hat{\underline{F}}_t^m)$. Let $\underline{F}_\infty^m := \bigvee_t \underline{F}_t^m$ and $\underline{F}_t := \bigvee_m \underline{F}_t^m$, $0 \leq t \leq \infty$. Let

$$X_t^m := E\{X|\underline{F}_t^m\}\ ,\quad 0 \leq t \leq \infty$$

be the estimate of agent $\underline{m}$ at t.

(iii) Let t_j^m, $j = 1,2,\ldots$ be an increasing sequence of $(\underline{F}_t^m)$-stopping times. At each time t_j^m, m sends the message $X_j^m := X_{t_j^m}^m$ to agents belong to the nonempty subset A_j^m. A_j^m is randomly selected, measurable with respect to $\underline{F}_{t_j^m}^m$.

It can be verified that the reception times r_j^{m0}, r_j^m are $(\underline{F}_t^m)$-stopping times. We do not insist that these be finite. However the transmission times are finite:

<u>Assumption 1</u>. $t_j^m < \infty$ and $\lim_j t_j^m = \infty$ a.s.

<u>Assumption 2</u>. Atmost one message transmission or reception occurs at any time. That is, for $j \neq k$, $t_j^m < t_{j+1}^m$ and $t_j^m \neq r_k^m$ a.s.
This assumption is inessential, but it relieves some notational burden. It could equally well be replaced by establishing some fixed priority ordering among messages received or transmitted simultaneously.

A <u>transception</u> at $\underline{m}$ is a transmission or reception of a message at $\underline{m}$. Let τ_n^m, $n = 1,2,..$ be the nth transception time. This is an a.s. finite, strictly increasing sequence of $(\underline{F}_t^m)$-stopping times such that

$$\bigcup_n \{\tau_n^m(\omega)\} = \bigcup_j \{t_j^m(\omega)\} \bigcup_k \{r_k^m(\omega) < \infty\} \quad .$$

Because of Assumption 1, $\lim_n \tau_n^m = \infty$ a.s.

Let $\underline{\underline{G}}_n^m$ be the σ-field generated by the first n transceptions i.e. by subsets of the form

$$\{Z_j^m \in B\} \cap \{r_j^m \leq \tau_n^m\}, \{A_j^m = \tilde{A}\} \cap \{X_j^m \in B\} \cap \{t_j^m \leq \tau_n^m\} ,$$

where $\tilde{A}$ is any subset of agents. $\underline{\underline{G}}_n^m$ does not explicitly include information contained in the signals from the environment - the Z_j^{m0}.

Let $r_k^m(p)$ be the time at which agent $\underline{m}$ receives the kth message from agent $\underline{p}$, and let $Z_k^m(p)$ be this message; set $r_k^m(p) = \infty$ if $\underline{m}$ receives from $\underline{p}$ fewer than k messages. The random variable $Z_k^m(p)$ is defined only on the event $\{r_k^m(p) < \infty\}$. Similarly, let $t_k^m(p)$ be the time at which $\underline{m}$ transmits a message to $\underline{p}$ for the kth time, and let $X_k^m(p)$ be this message; set $t_k^m(p) = \infty$ if $\underline{m}$ transmites to $\underline{p}$ fewer than k messages. Again, $X_k^m(p)$ is defined only on $\{t_k^m(p) < \infty\}$.

Finally, let $q_k^m = n$ if the kth message transmitted by $\underline{m}$ (to any other agent) occurs at the nth transception time i.e. if $t_k^m = \tau_n^m$. Then q_n^m is a $(\underline{\underline{G}}_n^m)$-stopping time, and $q_n^m < \infty$ a.s. by Assumption 1.

Assumption 3. Messages are subject to random but finite transmission delays. That is, for each $\underline{p}$, $\underline{m}$ and k,

$$\{t_k^m(p) < \infty\} = \{r_k^p(m) < \infty\} .$$

Moreover,

$$\{\lim_k t_k^m(p) = \infty\} = \{\lim_k r_k^p(m) = \infty\} .$$

This completes the specification of the model. Assumption 1 guarantees that each agent sends his estimate to some other agent infinitely often. As mentioned already, Assumption 2 is a technicality. In regard to Assumption 3 observe that if all agents have the same "absolute" time, then a message must be transmitted before it can be received i.e. $t_k^m(p) \leq r_k^p(m)$. However, we do not impose such a restriction, so that each agent may have his own "local" time. Nor is it required that the messages be received in the same order that they are transmitted, (think of the communication medium as the Post Office -- but Assumption 3 implies that messages cannot get lost or distorted, and messages are received out of order by only a finite amount.) The different local times cannot be completely arbitrary since all the transception times are defined on the same probability space, so that there is an implicit restriction. It would be an important extension of this model to specify each agent's operation solely in terms of his own local time, and make explicit the dependence between different local times imposed by message exchanges. For a discussion of some of the issues involved here see [3,4,5].

Let

$$\hat{G}_k^m := G^m_{q_k^m} \ , \ \hat{G}^m := \bigvee_k \hat{G}_k^m \ \text{and} \ \hat{X}_k^m := E\{X|\hat{G}_k^m\} \quad .$$

Recall that $X_t^m = E\{X|\underline{\underline{F}}_t^m\}$, $0 \leq t \leq \infty$ and $X_j^m = X_{t_j^m}$.

Theorem 1. (i) $(X_t^m, \underline{\underline{F}}_t^m)$, $0 \leq t \leq \infty$, is a martingale. (ii) $X_j^m = E\{X|\underline{\underline{F}}_{t_j^m}^m\}$ (iii) $\lim_{t\to\infty} X_t^m = \lim_{j\to\infty} X_j^m = \lim_{k\to\infty} \hat{X}_k^m = X_\infty^m$.

Proof. (i) follows from the fact that X_t^m, $0 \leq t < \infty$ is a uniformly integrable family. (ii) then follows from the optimal stopping theorem

[6, p. 133]. By the martingale convergence theorem [6, p. 131], $\lim X_t^m = X_\infty^m$ and hence (ii) implies $\lim X_j^m = X_\infty^m$. Finally, by optional sampling again, $\hat{X}_k^m = X_{q_k^m}^m$ and since $\lim_k q_k^m = \infty$ a.s., therefore $\lim \hat{X}_k^m = X_\infty^m$. ¤

Thus each agent's estimate converges. The interesting question is to determine when there is asymptotic agreement.

3. When is there asymptotic agreement

A random variable W is common knowledge for a family $\underline{\underline{H}}^1,\ldots,\underline{\underline{H}}^n$ of σ-fields if W is measurable with respect to each of them.

For any event F in $\underline{\underline{F}}$ let 1(F) denote its indicator function.

For any two agents m and p let S^{mp} be the event that m sends messages to p infinitely often (i.o.) Precisely,

$$S^{mp} := \{\omega | t_k^m(p)(\omega) < \infty,\ k = 1,2,..\}$$

$$= \{\omega | r_k^p(m)(\omega) < \infty,\ k = 1,2,..\}$$

by Assumption 3.

Lemma 1. Both $X_\infty^m\, 1(S^{mp})$ and $1(S^{mp})$ are common knowledge for $\underline{\underline{G}}_\infty^m$ and $\underline{\underline{G}}_\infty^p$. Moreover

$$X_\infty^m\, 1(S^{mp}) = E\{X | \underline{\underline{G}}_\infty^m \cap \underline{\underline{G}}_\infty^p\}\, 1(S^{mp}) \quad \text{a.s.}$$

Proof. Since

$$\{t_k^m(p) < \infty\} = \bigcup_n \{p \in A_j^m \text{ and } t_j^m \leq \tau_n^m \text{ for at least k distinct values of } j\},$$

it follows that $\{t_k^m(p) < \infty\}$, and hence S^{mp}, are in $\underline{\underline{G}}_\infty^m$. Similarly, $\{r_k^p(p) < \infty\}$ and S^{mp} are in $\underline{\underline{G}}_\infty^p$.

The kth message transmitted by $\underline{m}$ to $\underline{p}$, $X_k^m(p)$, is defined on $\{t_k^m(p) < \infty\}$, hence $X_k^m(p)\, 1\{t_k^m(p) < \infty\}$ is a well-defined random variable and by Theorem 1,

$$\lim_k X_k^m(p)\, 1\{t_k^m(p) < \infty\} = X_\infty^m\, 1(S^{mp}) \ . \tag{1}$$

Suppose the kth message transmitted by $\underline{m}$ to $\underline{p}$ is the ℓ_kth message received by $\underline{p}$ from $\underline{m}$, i.e.

$$X_k^m(p)\, 1\{t_k^m(p) < \infty\} = Z_{\ell_k}^p(m)\, 1\{t_k^m(p) < \infty\} \ . \tag{2}$$

ℓ_k is a random variable, and by Assumption 3, $\ell_k \to \infty$ as $k \to \infty$, hence

$$\lim_k Z_{\ell_k}^p(m)\, 1\{t_k^m(p) < \infty\} = \lim_\ell Z_\ell^p(m)\, 1(S^{mp})$$

which is $\underline{\underline{G}}_\infty^p$-measurable. But from (1), (2) this limit equals $X_\infty^m\, 1(S^{mp})$ so that the first assertion is true. The last assertion is now immediate since, by Theorem 1, $E\{X|\underline{\underline{G}}_\infty^m\} = X_\infty^m$. ¤

<u>Corollary 1</u>. $X_\infty^m\, 1(S) = X_\infty^p\, 1(S)$ a.s. where $S = S^{mp} \cap S^{pm}$.

<u>Proof</u>. Immediate from Lemma 1. ¤

The corollary says that if, on a particular sample path ω, $\underline{m}$ and $\underline{p}$ communicate with each other infinitely often, then they must agree asymptotically on ω. It seems reasonable to expect asymptotic agreement even when m and p communicate with each other i.o. through other agents rather than directly.

Definition. A sequence of not necessarily distinct agents $\underline{m}_1, \ldots, \underline{m}_{n+1} = \underline{m}_1$ forms a communicating ring for an event S in $\underline{F}$ if $S \subset S^{m_i m_{i+1}}$, $i = 1,\ldots,n$.

Theorem 2. Suppose that

(i) $\underline{m}_1, \ldots, \underline{m}_{n+1} = \underline{m}_1$ forms a communicating ring for S, and

(ii) $1(S)$ is common knowledge for $\underline{G}_\infty^{m_1}, \ldots, \underline{G}_\infty^{m_n}$.

Then the $X_\infty^{m_i}$ agree on S i.e. $X_\infty^{m_1} 1(S) = \ldots = X_\infty^{m_n} 1(S)$ a.s.

The following lemma is useful in the proof of the theorem.

Lemma 2. Let $W^1, \ldots, W^{n+1} = W^1$ be integrable random vectors and $\underline{H}^1, \ldots \underline{H}^n$ be σ-fields such that $W^i = E(W^{i+1}|\underline{H}^i)$, $i = 1,\ldots,n$. Then $W^1 = \ldots = W^n$ a.s.

Proof. Since the argument below applies to each component we may suppose that the W^i are scalars. Suppose first that the W^i are square integrable. Since conditional expectation is the best mean square estimate, therefore $W^i = E(W^{i+1}|\underline{H}^i)$ implies

$$E|W^{i+1}|^2 = E|W^i|^2 + E|\tilde{W}^i|^2 , \quad i = 1,\ldots,n$$

where $\tilde{W}^i := W^{i+1} - W^i$. Adding these relations and recalling that $W^{n+1} = W^1$ gives

$$0 = E|\tilde{W}^1|^2 + \ldots + E|\tilde{W}^n|^2$$

and so $W^1 = \ldots = W^n$. Thus the assertion holds for square integrable random variables.

Next, for any number x, let $W_x^i := \min\{W^i, x\}$. By Jensen's inequality, $W^i = E(W^{i+1}|\underline{H}^i)$ implies

$$W_x^i \geq E(W_x^{i+1}|\underline{H}^i) , \tag{3}$$

and since from (3) we get $EW_x^1 \geq EW_x^2 \geq .. \geq EW_x^{n+1} = EW_x^1$, so (3) holds with equality. Therefore, for $x > y$,

$$W_x^i - W_y^i = E(W_x^{i+1} - W_y^{i+1} | \underline{\underline{H}}^i)$$

and since $W_x^i - W_y^i$ is bounded, hence square integrable, therefore $W_x^1 - W_y^1 = .. = W_x^n - W_y^n$. The result now follows by letting $x \to \infty$, $y \to -\infty$.

¤

<u>Proof of Theorem 2</u>. By Lemma 1

$$\begin{aligned} X_\infty^{m_i} 1(S^{m_i m_{i+1}}) &= E(X | \underline{\underline{G}}_\infty^{m_i} \cap \underline{\underline{G}}_\infty^{m_{i+1}}) \, 1(S^{m_i m_{i+1}}) \\ &= E(X_\infty^{m_{i+1}} | \underline{\underline{G}}_\infty^{m_i} \cap \underline{\underline{G}}_\infty^{m_{i+1}}) \, 1(S^{m_i m_{i+1}}) \\ &= E\{X_\infty^{m_{i+1}} \, 1(S^{m_i m_{i+1}}) | \underline{\underline{G}}_\infty^{m_i} \cap \underline{\underline{G}}_\infty^{m_{i+1}}\} \quad . \end{aligned} \tag{4}$$

By hypothesis (ii), $S \in \underline{\underline{G}}_\infty^{m_i}$ and $S \subset S^{m_i m_{i+1}}$. Hence, multiplying both sides of (4) by 1(S) gives

$$W^i = E(W^{i+1} | \underline{\underline{H}}^i)$$

where $W^i := X_\infty^{m_i} 1(S)$ and $\underline{\underline{H}}^i := \underline{\underline{G}}_\infty^{m_i} \cap \underline{\underline{G}}_\infty^{m_{i+1}}$. By Lemma 2, $W^1 = .. = W^n$, and the assertion follows.

¤

<u>Corollary 2</u>. Under the hypotheses of Theorem 2

$$X_\infty^{m_i} 1(S) = E\{X | \underline{\underline{G}}_\infty^{m_1} \cap .. \cap \underline{\underline{G}}_\infty^{m_n}\} \, 1(S) \; .$$

<u>Proof</u>. By (4)

$$X_\infty^{m_i} 1(S) = E\{X \, 1(S) | \underline{\underline{G}}_\infty^{m_i} \cap \underline{\underline{G}}_\infty^{m_{i+1}}\} \; . \tag{5}$$

By Theorem 2, $X_\infty^{m_i} 1(S) = X_\infty^{m_j} 1(S)$ and so $X_\infty^{m_i} 1(S)$ is common knowledge for

$\underline{\underline{G}}_\infty^{m_1}, \ldots, \underline{\underline{G}}_\infty^{m_n}$. Hence taking conditional expectation with respect to $\underline{\underline{G}}_\infty^{m_1} \cap .. \cap \underline{\underline{G}}_\infty^{m_n}$ in (5) gives

$$X_\infty^{m_i} 1(S) = E\{X\, 1(S) | \underline{\underline{G}}_\infty^{m_1} \cap .. \cap \underline{\underline{G}}_\infty^{m_n}\}$$

$$= E\{X | \underline{\underline{G}}_\infty^{m_1} \cap .. \cap \underline{\underline{G}}_\infty^{m_n}\}\, 1(S) \;,$$

since by hypothesis $S \in \underline{\underline{G}}_\infty^{m_i}$, all i. ¤

Corollary 2. Suppose that with probability one, all agents are members of the same communicating ring. Then the estimate of each agent converges to $E\{X | \underline{\underline{G}}_\infty^1 \cap .. \cap \underline{\underline{G}}_\infty^M\}$.

Proof. Immediate from Corollary 2 since $1(\Omega)$ is certainly common knowledge for all. ¤

Hypothesis (i) of Theorem 2 says that, over the event S, every pair of agents in the communicating ring send messages i.o. to each other either directly or through other agents in the ring. Hypothesis (ii) says moreover that, over the event S, each agent in the ring knows that he is a member of the ring. This second hypothesis is not implied by the first. For example, if for a sample path ω, $\underline{m}_1$ sends i.o. his estimate to $\underline{m}_2$ and $\underline{m}_3$ sends i.o. his estimate to $\underline{m}_1$, $\underline{m}_1$ may not be able to infer whether or not $\underline{m}_2$ sends i.o. his estimate to $\underline{m}_3$ (and then m_1, m_2, m_3, m_1 forms a ring for ω), even though $\underline{m}_3$'s message to $\underline{m}_1$ will depend upon whether or not and how often he hears from $\underline{m}_2$.

It may be conjectured, however, that the fact that $\underline{m}_1, \underline{m}_2, \ldots, \underline{m}_{n+1}$ form a ring is sufficient to guarantee asymptotic agreement i.e. that hypothesis (ii) is unnecessary in Theorem 2. We now give a counterexample to this conjecture.

Example 1. Take $\Omega = \{a,..,g\}$, $\underline{\underline{F}} = 2^{\Omega}$, and let P be uniformly distributed on Ω. There are three agents labelled 1, 2, 3. The table below gives the random variables X, X^1, X^2, X^3. Thus, for instance, for $\omega = a$, the random variables take the values indicated by column a, so $X(a) = 1$, $X^1(a) = 0.5$, $X^2(a) = 1$

ω	a	b	c	d	e	f	g
X	1	0	-2	1	0	2	1
X^1	.5	.5	-2	.5	.5	1.5	1.5
X^2	1	0	0	1	0	0	1
X^3	1	-1	-1	1	1	1	1

and $X^3(a) = 1$. For i = 1, 2, 3 let $\underline{\underline{G}}^i$ be the subfield of $\underline{\underline{F}}$ generated by the partition $\underline{\underline{P}}^i$ where

$$\underline{\underline{P}}^1 = \{a,b\}, \{d,e\}, \{f,g\}, \{c\}$$

$$\underline{\underline{P}}^2 = \{a,d\}, \{b,e\}, \{c,f\}, \{g\}$$

$$\underline{\underline{P}}^3 = \{d,g\}, \{e,f\}, \{b,c\}, \{a\} .$$

Observe first that $X^i = E(X|\underline{\underline{G}}^i)$. Define the events $S^{12} = \{a,b,d,e\}$, $S^{23} = \{b,e,f,c\}$, $S^{31} = \{d,e,f,g\}$, and check that $X^i\ 1(S^{i,i+1})$ and $1(S^{i,i+1})$ are common knowledge for $\underline{\underline{G}}^i$, $\underline{\underline{G}}^{i+1}$ (Here and below i+1 = 4 refers to agent 1.) Also check that $X^i\ 1(S^{i,i+1}) = E\{X^{i+1}|\underline{\underline{G}}^i \cap \underline{\underline{G}}^{i+1}\}\ 1(S^{i,i+1})$. Suppose now that agent i receives signal X^i from the environment, and when $\omega \in S^{i,i+1}$ he transmits X^i to agent i+1. It is easy to check that under these circumstances X^i and $\underline{\underline{G}}^i$ remain unchanged, so that in terms of the previous notation $\underline{\underline{G}}^i = \underline{\underline{G}}^i_\infty$, and $X^i = X^i_\infty$. Now $S^{12} \cap S^{23} \cap S^{31} = \{e\}$, so all agents form a communicating ring over {e}. However, $\{e\} \notin \underline{\underline{G}}^i$ for every i, i.e. none of the agents can know if they do form a ring, so that hypothesis (ii) of Theorem 2 does not hold. And indeed $X^1(e)$, $X^2(e)$ and $X^3(e)$ are all different, so that the conclusion of Theorem 2 does not hold either.

3. Properties of the asymptotic estimate

Throughout this section it is assumed that with probability one, all agents form a communicating ring. By Corollary 3 then, all estimates agree asymptotically and equal $\hat{X}$, where

$$\hat{X} := E(X|\underline{\underline{G}}_\infty^1 \cap .. \cap \underline{\underline{G}}_\infty^M) = E(X|\underline{\underline{F}}_\infty^m) = X_\infty^m \quad . \tag{6}$$

(In case all agents form a communicating ring over S and 1(S) is common knowledge for all $\underline{\underline{G}}_\infty^m$, then the discussion below continues to apply provided X is replaced by X 1(S).)

Now if the agents shared their "raw data" rather than their estimates, then their estimates would converge asymptotically to the <u>full information estimate</u> X^*,

$$X^* := E(X|\underline{\underline{F}}_\infty^1 \vee .. \vee \underline{\underline{F}}_\infty^M) \ . \tag{9}$$

Since $\hat{X} = E(X^*|\underline{\underline{G}}_\infty^1 \cap .. \cap \underline{\underline{G}}_\infty^M)$, therefore X^* is a better estimate than $\hat{X}$. The next example shown that X^* can be strictly better.

<u>Example 2</u>. Let $\Omega = [0,1]^2$, $\underline{\underline{F}}$ the Lebesgue field and P the Lebesgue measure. Agents 1 and 2 try to estimate the random variable X := 1(A) where A is the hatched region in Figure 1. At t = 0, agent 1 observes 1(B), and agent 2 observes 1(C). At t = 1, 3, 5, ... agent 1 transmits his estimate X_t^1 to 2 and at t = 2, 4, 6, .. agent 2 transmits his estimate X_t^2 to 1. Since 1(A), 1(B), 1(C) are pairwise independent, $X_t^1 = X_t^2 = \hat{X} = EX = 0.5$ a.s., whereas if the agents shared their observations the

resulting estimate X^* would take the values 0.25 and 0.75 each with probability 0.5. Thus $\hat{X} \neq X^*$.

Thus the raw data contains information beyond what is common knowledge and which can be used to provide a better estimate of X. However the additional information available to any single agent must be uncorrelated with X given the common knowledge $\underline{\underline{G}}_\infty := \underline{\underline{G}}^1_\infty \cap .. \cap \underline{\underline{G}}^M_\infty$.

<u>Lemma 3</u>. Let $F \in \underline{\underline{F}}^m_\infty$. Then $E\{X\ 1(F)|\underline{\underline{G}}_\infty\} = \hat{X}\ E\{1(F)|\underline{\underline{G}}_\infty\}$.

<u>Proof</u>. Using (8),

$$E\{X\ 1(F)|\underline{\underline{G}}_\infty\} = E\{E\{X\ 1(F)|\underline{\underline{F}}^m_\infty\}|\underline{\underline{G}}_\infty\} = \hat{X}\ E\{1(F)|\underline{\underline{G}}_\infty\}\ . \qquad \square$$

The lemma shows that in the Gaussian case $\hat{X} = X^*$. More precisely, by the <u>Gaussian case</u> we mean that (i) X and the signals received from the environment -- the Z^{m0}_j -- are all jointly Gaussian and (ii) the transmission and reception times as well as the sets A^m_j are all deterministic. Then $\underline{\underline{F}}^m_\infty$ and $\underline{\underline{G}}_\infty$ are generated by sequences of Gaussian random variables, and so by Lemma 3, for each m, X and $\underline{\underline{F}}^m_\infty$ are independent given $\underline{\underline{G}}_\infty$. But then X and $\bigvee_m \underline{\underline{F}}^m_\infty$ are also independent given G , which proves the next result.

<u>Lemma 4</u>. In the Gaussian case $\hat{X} = X^*$.

Now suppose that each agent receives exactly one signal from the environment at t = 0 and that there is no transmission delay. The next example shows that the common asymptotic estimate can depend upon the sequence in which messages are exchanged.

<u>Example 3</u>. Let $\Omega = [0,2] \times [0,3]$, $\underline{\underline{F}}$ the Lebesgue field and P the Lebesgue measure (normalized to give $P(\Omega) = 1$). Agents 1 and 2 try to estimate 1(A), where A is shown in Figure 2. At t = 0, agent 1 observes

$1(B_1)$ and agent 2 observes $\{1(C_1), 1(C_2)\}$. At $t = 1, 3, 5,..$ agent 1 sends his estimate X_t^1 to 2 and at $t = 2, 4, 6,..$ the latter sends his estimate X_t^2 to agent 1.

Now $X_1^1(\omega) := E\{1(A)|1(B_1)\}(\omega) = \frac{1}{2}$ or $\frac{1}{3}$ according as $\omega \in B_1$ or $\omega \in B_2$. Hence $\sigma(X_1^1)$ is generated by the partition $\{B_1, B_2\}$. After receiving X_1^1 therefore, agent 2's information field is generated by the partition $\{B_i \cap C_j;\ i = 1,2,\ j = 1,2,3\}$. Since this is also the field generated by all the observations $1(B_1)$, $1(C_1)$, $1(C_2)$, therefore $X_2^2 = E\{X|1(B_1), 1(C_1), 1(C_2)\} =: \hat{X}$ is the asymptotic estimate. Note that $\hat{X}(\omega) = \frac{1}{4}$ if $\omega \in B_1 \cap C_3$.

Suppose now agent 2 sends his estimate X_t^2 at $t = 1, 3, 5,..$ and agent 1 sends X_t^1 at $t = 2, 4, 6,..$. Then $X_1^2(\omega) =: E\{X|1(C_1), k(C_2)\}(\omega) = \frac{1}{4}$ or $\frac{1}{2}$ according as $\omega \in C_1$ or $\omega \in C_2 \cup C_3$. Hence $\sigma(X_1^2)$ is generated by the partition $\{C_1, C_2 \cup C_3\}$. After receiving X_1^2 therefore, agent 1's information field is generated by the partition $\{B_i \cap C_1, B_i \cap (C_2 \cup C_3);\ i = 1,2\}$. Hence $X_2^1(\omega) := E\{X|1(B_1), X_1^2\}(\omega) = 0$ or $\frac{1}{2}$ according as $\omega \in B_2 \cap C_1$ or $\omega \notin B_2 \cap C_1$, and so $\sigma(X_2^1)$ is generated by the partition $\{B_2 \cap C_1, (B_2 \cap C_1)^c\}$. After receiving X_2^1 therefore, agent 2's information field in generated by the partition $\{B_1 \cap C_1, B_2 \cap C_1, C_2 \cup C_3\}$. But then $X_3^2 = X_2^1 =: \tilde{X}$ must also be the asymptotic estimate. But note that $\tilde{X} \neq \hat{X}$, since $\tilde{X}(\omega) = \frac{1}{2}$ and $\hat{X}(\omega) = \frac{1}{4}$ for $\omega \in B_1 \cap C_3$.

4. Distributed detection

The setup is almost the same as in section 2 except that the message X_t^m transmitted by any agent m at time t is different.

The space of events $(\Omega, \underline{F})$ admits two probability distributions P_1 and P and the agents need to decide which of these is "true." Agent m's decision at time t is based on the conditional likelihood ratio,

$$X_t^m := E(\frac{dP_1}{dP} \Big| \underline{\underline{F}}_t^m) . \qquad (10)$$

The messages consist of these ratios.

For any σ-field $\underline{\underline{G}} \subset \underline{\underline{F}}$ say that $P_1 << P|\underline{\underline{G}}$ if, restricted to $\underline{\underline{G}}$, P_1 is absolutely continuous with respect to P, that is for $G \in \underline{\underline{G}}$, $P(G) = 0$ implies $P_1(G) = 0$.

Recall that $\underline{\underline{F}}_t = \bigvee_m \underline{\underline{F}}_t^m$.

Assumption 4. For each t, $P_1 << P|\underline{\underline{F}}_t$.

Then $X_t := E(\frac{dP_1}{dP} \Big| \underline{\underline{F}}_t)$ is well-defined when interpreted to mean the Radon-Nikodym derivative of P_1 with respect to P restricted to $\underline{\underline{F}}_t$. X_t^m in (10) is similary interpreted. Observe that if $P_1 << P|\underline{\underline{F}}_\infty$, then $X_t^m = E(X|\underline{\underline{F}}_t^m)$ where $X := \frac{dP_1}{dP}$, and so the problem reduces to the one considered already. Hence the discussion below is of interest only when P_1 is not absolutely continuous with respect to P and Assumption 4 holds. This is the usual situation in detection or system identification problems.

Here is the result corresponding to Theorem 1.

Theroem 3. (i) $(X_t^m, \underline{\underline{F}}_t^m)$, $0 \le t < \infty$ is a martingale, (ii) the limit $X_\infty^m := \lim_t X_t^m$ a.s. exists. (iii) $(X_t^m, \underline{\underline{F}}_t^m)$, $0 \le t \le \infty$ is a supermartingale.

Proof. The first assertion follows from (10). Also X_t^m is obviously nonnegative and so (ii), (iii) follow from [6, p. 131]. ¤

We now investigate asymptotic agreement. The same argument as in the proof of Lemma 1 shown that $X_\infty^m 1(S^{mp})$ and $1(S^{mp})$ are common knowledge for $\underline{\underline{G}}_\infty^m$, $\underline{\underline{G}}_\infty^p$. If P_1 is not absolutely continuous with respect to P restricted to $\underline{\underline{F}}_\infty$, then the last assertion in Lemma 1 does not hold, and so the proof of Theorem 2 no longer applies. Indeed we are unable to prove asymptotic agreement without some additional restirctions. Several alternatives are possible, and we present one of these.

Assumption 5. The transmission delay between any two agents is fixed.

This assumption implies that messages are received in the order in which they are sent. Let $\underline{\underline{G}}_t^p(m,k)$ be the σ-field generated by the first k messages received by $\underline{p}$ from $\underline{m}$ before t. $\underline{\underline{G}}_t^p(m,k)$ is generated by subsets of the form

$$\{Z_\ell^p(m) \in B,\ r_\ell^p(m) \leq s\}\ ,\quad \ell \leq k,\ s \leq t\ . \tag{11}$$

If Δ is the delay experienced by messages sent by $\underline{m}$ to $\underline{p}$, then $\underline{\underline{G}}_t^p(m,k)$ is also generated by the first k messages sent by $\underline{m}$ to $\underline{p}$ before t-Δ i.e. by subsets of the form

$$\{X_\ell^m(p) \in B,\ t_\ell^p(m) \leq s\},\quad \ell \leq k,\ s \leq t-\Delta\ . \tag{12}$$

Lemma 5. $E\{X_t^p|\underline{\underline{G}}_t^p(m,k)\}\ 1(r_k^p(m) \leq t) = X^m_{t_k^m(p)}\ 1(t_k^m(p) \leq t-\Delta)\ .$ (13)

Proof. From (10), (11) it follows that the random variable on the left equals

$$E\{X_t|\underline{\underline{G}}_t^p(m,k)\}\ 1(t_k^m(p) \leq t-\Delta)\quad ,$$

and by Theorem 3 (i) and (12), this equals the variable on the right. ¤

Letting $t \to \infty$ in (13), and using Theorem 3, gives

$$E\{X_\infty^p|\underline{\underline{G}}_\infty^p(m,k)\}\ 1(t_k^m(p) < \infty) \leq X^m_{t_k^m(p)}\ 1(t_k^m(p) < \infty)\quad .$$

Letting $k \to \infty$, and again using Theorem 3, gives

$$E\{X_\infty^p|\underline{\underline{G}}_\infty^p(m)\}\ 1(S^{mp}) \leq X_\infty^m\ 1(S^{mp})\ , \tag{14}$$

where $\underline{\underline{G}}^p_\infty(m) := \bigvee_k \underline{\underline{G}}^p_\infty(m,k)$ is the σ-field generated by all messages received by $\underline{p}$ from $\underline{m}$.

Theorem 4. Suppose that

(i) $\underline{m}_1,\ldots,\underline{m}_{n+1} = \underline{m}_1$ forms a communication ring for S, and

(ii) $1(S)$ is common knowledge for $\underline{\underline{G}}^{m_{i+1}}(m_i)$, $i = 1,\ldots,n$.

Then the $X_\infty^{m_i}$ agree on S.

Proof. Clearly $S^{mp} \subset \underline{\underline{G}}^p_\infty(m)$. Hence the hypothesis and (14) give

$$E\{X^{m_{i+1}}\, 1(S) \,|\, \underline{\underline{G}}_\infty^{m_{i+1}}(m_i)\} \le X^{m_i}\, 1(S) .$$

Since $\underline{m}_{n+1} = \underline{m}_1$, there must be equality. The assertion now follows from Lemma 2. ¤

5. Concluding remarks

A distributed estimation system consists of several local sensors and processors, each sensor feeds "raw" data to one or more processors, and the processors in turn are linked by a communication network. One of the decisions that must be taken in designing such a system is to determine how much "preprocessing" of data should be done by each local processor before it communicates to its neighbors. Underlying this tradeoff is the presumption that the greater the amount of preprocessing (more precisely, the greater the raw data is stripped of irrelevant components), the less is the need for inter-processor communication capacity.

Intelligent use of processed data requires a knowledge of how the original data relates to the processed data. In terms of the model of Section 2, if $\underline{p}$ receives from $\underline{m}$ the latter's current estimate X_t^m, the proper use of this statistic requires that $\underline{p}$ should know the joint distribution of X_t^m and X, a knowledge which would normally be derived from the joint distribution of the messages that m has received. Since the relation between raw and processed data is often complex, use of the latter will then entail a lot of computation. These computational tasks, moreover, are very contingent on the manner in which processing is done locally, so that a change in this processing at one location will propagate changes everywhere.

These remarks concerned with the implementation of distributed implementation and related considerations (see Wong [7]), suggest that preprocessing of data should be limited to the simplest possible operations. The results presented here reinforce such a conclusion. In the first place it is clear that an estimate based on processed data is inferior to the full information estimate (cf. Section 3). A more subtle observation stems from the fact that the former depends upon the ordering of inter-processor messages (cf. Example 3), thereby introducing an additional complication in the design of distributed estimation systems.

The specific problems discussed in this paper raise two additional questions. The first which concerns formulating distributed estimation problems solely in terms of local description was already stated in Section 2. The second concerns the role of a priori information. Our discussion stipulates a single probability space. This presupposes that all processor-agents have the same "view of the world," in particular the same prior probability. There are many situations (horse races, the stockmarket?) where agents with different beliefs exchange information. How can we model the process of expectation formation in such systems?

REFERENCES

1. R. J. Aumann, "Agreeing to disagree," The Annals of Statistics 4 (6), 1976, pp. 1236-1239.

2. J. D. Geanakoplos and H. M. Polemarchakis, "We can't disagree forever," Institute for Mathematical Studies in the Social Sciences, Technical Report No. 277, Stanford University, Stanford, California, December 1978.

3. C. Courcoubetis and P. Varaiya, "A preliminary model for distributed algorithms," presented at 2nd IFAC Symposium on Large Scale Systems: Theory and Applications, Toulouse, France, June 24-26, 1980.

4. L. Lamport, "Time, clocks and the ordering of events in a distributed system, Communications of the ACM 21 (7), 1978, pp. 558-565.

5. H. S. Witsenhausen, "Some remarks on the concept of state," in Y. C. Ho and S. K. Mitter (eds.), Directions in Large-Scale Systems, New York: Plenum Press, 1976.

6. P-A. Meyer, Probabilités et potentiel, Paris: Hermann, 1966.

7. E. Wong, "Incomplete information in database systems," 19th IEEE Conference on Decision and Control, Albuquerque, New Mexico, December 8-12, 1980.

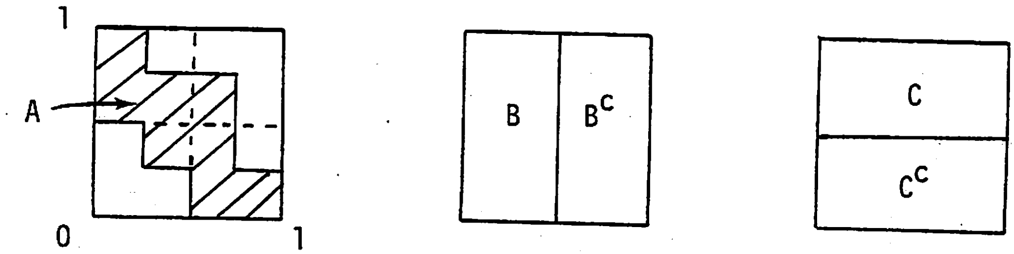

Fig. 1. Parameters for Example 2.

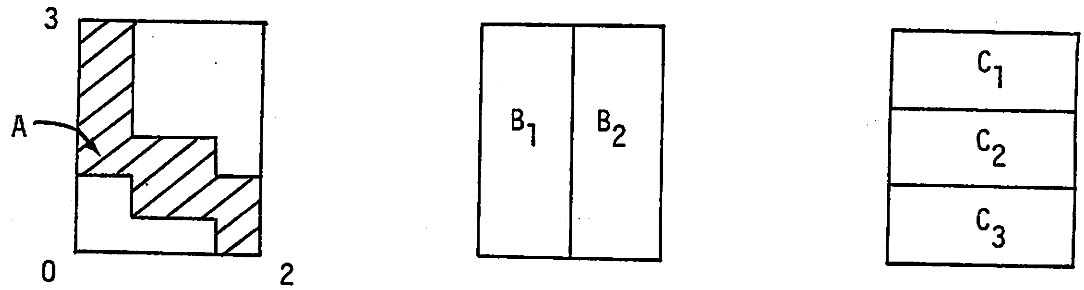

Fig. 2. Parameters for Example 3.