Conditional Expectations and Pouring Water from Full Cups to Empty.

Stanislav A. Molchanov, Isaac M. Sonin

Department of Mathematics and Statistics
University of North Carolina at Charlotte, USA
ЦЭМИ РАН, Russia

http://...Isaac Sonin in Google

Moscow, December 2015

Part 1. Consensus algorithms and the Decomposition-Separation Theorem

Традиционно *консенсус* (К) определяется как согласие по некоторому вопросу, идея или мнение, с которым согласны участники некоторой группы. В наши дни это определение устарело. Как показали многочисленные исследования последних десятилетий, многие системы «частиц», «агентов» проявляют кооперативное поведение ведущее к К-у. Примеры такого рода широко распространены в биологии, роботехнике, социологии, политологии, физике, механике, информатике и т.п. *Консенсусные алгоритмы* это алгоритмы ведущие к К. Наиболее изученные из них основаны на линейных преобразованиях и стохастических матрицах. 30 лет тому назад я придал заключительную форму теореме названной the Decompostion-Separation theorem и начатой в 30х годах А. Колмогоровым. Как утверждается в одной из текущих публикаций "D-S Theorem relates to and generalizes most of the existing results in the literature on convergence properties of linear consensus algorithms".

Consensus

Consensus in Russian ? West vs Soviet Union, Russia

Стадо слонов Кто их ведет ?

Стая птиц ? murmuration

Саранча Пушкин

some bacteria exhibit cooperative motion

trees
В мастерской часовщика в средние века !

Lasers and Masers. A laser differs from other sources of light in that it emits light *coherently*.

Наше сердце, наш мозг. electr activity
Human Personality

Acemoglu Central Q-n in Social Sciences ...A Social Networks

Local Automata

Consensus. Terms

A multi-agent system, in the most general sense, is a network of multiple interacting *agents*. Each agent is assumed to hold a *state* regarding a certain quantity of interest. Depending on the context, states may be referred to as *opinions, values, beliefs, positions, velocities*, etc. States of agents are updated based on an algorithm or protocol which is an interaction rule specifying the interaction between each agent and its neighbors. Global *consensus*, or simply consensus, in the system is defined as *convergence of all states to a common value* over time. Among all update algorithms in multi-agent systems, *distributed averaging algorithms* are of great importance and have been discussed the most in the literature. Such algorithms impose that the state of each agent is updated according to a convex combination of the current states of its neighbors and its own.

Cucker & S Smale (2007) positions, velocities, nonlinear interaction phase transitions.

Consensus. More areas of application

In *biology* - behavior of bird flocks, fish schools, humans etc

In *robotics and control*, consensus problems arise in relation to coordination objectives and cooperation of mobile agents (e.g.,robots and sensors)

In *economics*, seeking an agreement on a common belief in a price system is another example of consensus. Gas prices in Ch-te

In *sociology*, the emergence of a common language in primitive societies
In *social networks*, consensus algorithms can shed light on the dynamics of opinion formation.

In *computer science* - networks; management science
In a multi-agent system, it is possible that agents separate into several clusters such that consensus occurs within each cluster. In this case, *multiple consensus* is said to have occurred.

Gossip algorithms. In a gossip models the frequency of information exchange is controlled by an internal clock ticking according to a timing model. In each step, each agent transmits its information (state) to another agent which is chosen randomly.

Consensus. Questions

Consider a system composed of N agents that are labeled by numbers $1, ..., N$. Let $f_n(i)$ be the scalar state of agent i at time n. *Distributed averaging algorithms* can be defined in both continuous and discrete times. A general discrete time *averaging algorithm* is defined by

$$f_{n+1} = P_n f_n, f_3 = P_2 P_1 P_0 f_0, \tag{1}$$

where f_n is the **column** vector of states at each time instant n. *Consensus* is now defined by the convergence of vector f_n to a vector with equal components as n.... *Multiple consensus* is also defined as the existence of a limit for $f_n(i)$ for each agent i as time grows large. The limits may differ for different agents.

The following two fundamental questions regarding the issue of consensus:

Q 1. Under what conditions on the underlying chain of the system, consensus or multiple consensus is guaranteed *irrespective* of the time and values that states are initialized ?

Consensus. Questions

Q 2. For a general underlying chain, having fixed the initial time, what is the *set of initial conditions* resulting in the occurrence of consensus in the system ?

Q 1 is equivalent to a property of the underlying chain called *ergodicity*. multiple consensus is equivalent to another property of underlying chain called *class-ergodicity.* Chain is class-ergodic if the limit matrix exists, but in general possibly with distinct rows.

Q 3. the question arises as to whether it is possible, for a limited number of *key agents*, to set their initial opinion/parameter assessment, in such a way that the (exogenously evolving) network converges to a global consensus. Such an issue is important in negotiations, or even the possible shaping or manipulation of public opinion by clever campaigning. notion of *éminence grise coalition.* Gray Cardinals.

For which MCs the consensus occurs ? It depends.

Markov Chains

The classical Kolmogorov-Doeblin results describing the asymptotic behavior of MCs can be found in most advanced books on probability theory.
According to these results the state space M can be *decomposed* into the set of nonessential states and the classes of essential communicating states. Furthermore, the following are true:

(**A**) With probability one, each trajectory of a MC Z from U_0 will reach one of these classes and never leave it.
Each class S can be decomposed into cyclical subclasses. If the number of subclasses is equal to one (an aperiodic class), then

(**B**) every MC Z from U_0 has a mixing property inside such a class, i.e. there exists a limit distribution π which does not depend on the initial distribution μ and such that π is invariant with respect to the matrix P.

What is a *nonhomogeneous MC* ? Replace matrix P by a sequence P_n.

Nonhomogeneous Markov Chains as Colored Flows

The following simple physical model and physical interpretation of the DS Theorem was introduced by I. Sonin in..1987. Given a sequence (M_n), let M_n represent a set of “cups” containing a “liquid ” - *tea*, schnapps, vodka etc. A cup $i \in M_n$ is characterized at moment n by a *volume* of liquid in this cup, $m_n(i)$. The matrix P_n describes the redistribution of liquid from the cups M_n to the (initially empty) cups M_{n+1} at the time of the n-th transition, i.e. $p_n(i,j)$ is the proportion of liquid transferred from cup i to cup j. The sequence $(m_n), m_n = (m_n(i), i \in M_n), n \in \mathbf{N}$, satisfies the relations

$$m_{n+1} = m_n P_n, m_3 = m_0 P_0 P_1 P_2, \tag{2}$$

Where m_n is a stochastic *row* vector. Let us assume additionally that each cup contains some material (substance, color) and let us denote $\alpha_n(i), 0 \leq \alpha \leq 1,$ a "*concentration*" of this material at cup i at moment n. The sequence $(m_n, \alpha_n) = (m_n(i), \alpha_n(i)), i \in M_n, n \in \mathbf{N})$, for the sake of brevity is called (discrete) *colored flow.*

Concentrations are Martingales in Reverse Time

Concentrations obviously satisfy the relations

$$\alpha_{n+1}(j) = \sum_i m_n(i)\alpha_n(i)p_n(i,j)/m_{n+1}(j). \tag{3}$$

Note that we can replace the notion of concentration by *temperature* since it follows the same formula. One more interpretation...
A random sequence (Y_n) specified by

$$Y_n = \alpha_n(Z_n), n \in \mathbf{N}, \tag{4}$$

where $\alpha_n(i)$ *s*' are given by ..., is a *(sub)martingale in reverse time*. This simple fact is the bridge between the DS theorem and the Theorem about the existence of *barriers*.

Doob's Lemma and its modification

One of the most remarkable and widely used results in the theory of stochastic processes is the theorem of Doob about the existence of the limits of trajectories of bounded (sub)martingale when time tends to infinity. This theorem is based on Doob's upcrossing lemma.

Doob's Lemma. *If* $Y = (Y_n)$ *is a bounded (sub)martingale then the expected number of intersections of every fixed interval* $a, b)$ *by the trajectories of* Y *is finite on the infinite time interval*.

The width of the interval $(b - a)$ is in the denominator of the corresponding estimate so Doob's lemma does not imply for example that inside the interval there exists a*level* such that the expected number of intersections of this level is finite.

If (Y_n) takes values in (M_n), then Doob's lemma can be substantially strengthened. Let us call a nonrandom sequence (d_n) a *barrier* for the random sequence $Y = (Y_n)$ if the *expected number* of intersections of (d_n) by the trajectories of X is finite, i.e...

Theorem in Sonin (1987) about the existence of barriers for processes with finite variation and which take only a bounded number of values implies the

DS Theorem. The elementary (deterministic) formulation

Let a sequence of disjoint sets (M_n), *satisfying condition* $|M_n| \leq N$ *and a sequence of stochastic matrices* (P_n) *be given. Then there an integer* $c, 1 \leq c \leq N$, *and there exists a decomposition of the sequence* (M_n) *into disjoint jets* $J^0, J^1, ..., J^c, J^k = (J^k_n)$, *such that for any colored flow* (m_n, α_n, O_n)

(**a**) *the stabilization of volume and concentration take place inside of any jet* J^k, $k = 1, ..., c$,

i.e. $\lim_{n\to\infty} \sum_{i \in J^k_n} m(i) = m^k_*$; $\lim_{n\to\infty} \alpha(i_n) = \alpha^k_*, i_n \in J^k_n$;

the concentration in jet J^0 *may oscillate; the total volume in this jet tends to zero, i.e.* $\lim_{n\to\infty} \sum_{i \in J^0_n} m(i) = 0$;

(**b**) *the total amount of liquid transferred between any two different jets is finite on the infinite time interval, i.e.* $V(J^k, J^s | m) < \infty, s \neq k$.

(**c**) *this decomposition is unique up to jets* (J_n) *such that for any flow* (m_n) *the relation* $\lim_n m_n(J_n) = 0$ *holds and the total amount of liquid transferred between* (J_n) *and* $(M_n \backslash J_n)$ *is finite.*

Part 2. Independence

A few years ago I wrote a small note... simple random experiment: first we flip a *fair* coin and then we toss a *fair* die. The sample space consists of 12 outcomes each having a probability of $1/12$. First, I answered two simple questions.

Question 1. How many different *pairs* (A, B) of independent events are there ? **Question 2.** How many different tuples...

Let the numbers K_1 and K_2 be the answers to these questions. The first number K_1 is equal to a rather strange looking number $888, 888.$

...most of these pairs and tuples are isomorphic and can be obtained

Let us suppose now that a coin and a die are (slightly) *biased*. Then, it is easy to check that for almost all biased coins and dice the number K_1 is reduced to the more "normal" looking number $124 = (2^2 - 2) \cdot (2^6 - 2).$ These 124 pairs are "stable", i.e. they are not affected by the changes in probabilities. This sample space, with 12 equally likely points can be represented also as a product of 3-die (a die with three sides) and 4-die, or as a product of two coins and a 3-die. The the number of stable pairs will be even smaller, $(2^3 - 2) \cdot (2^4 - 2) = 84$ and $2 \cdot 2 \cdot (2^3 - 2) = 24$. Teor. Ver.

Two Theorems. Open Problems

...a couple of almost trivial statements.

Proposition 1 *For any $N = |S|$ there are models without independent events.* Let us call such models *indecomposable*.

Proposition 2 *For any $N \geq 4$ there are models with at least one pair of ind. events.*

Theorem 1. *Any finite model can be uniquely decomposed into a product of indecomposable models.*

If $12 = 3 * 4$ but $4 \neq 2 * 2$ then $12 \neq 2 * 6$. The title has the following meaning: .... The equivalent statement-If $12 = 3 * 4$ and $12 = 2 * 6$ then$4 = 2 * 2$, $6 = 2 * 3$ and the 2s are the same.

This Theorem is wrong ! A simple counterexample

Example Let $N = 6$ and the prb mass function is given as follows: $\{1, 2, 4, 8, 16, 32\} * \frac{1}{63}$. Then it has two *distinct* factorizations.

Theorem 2. *In every model where there are dependent events which are partially independent, this partial independence is unstable.*

This Theorem confirms what Feller said in his famous book, after the definition of independence. "There is no practical cases of indep. events which are pairwise indep. but not indep."

Part 3 (joint with S. Molchanov). Cups again

Suppose we have $2n$ cups, n on the "left"and n on the "right each having volume $1/n$. The left cups are filled with water and the right cups are empty. Each of the cups is connected by pipes with all other cups and these pipes can open or closed at each moment of discrete time and if pipes between some cups are open this allows the water in them to come to equilibrium. Or left cups contain red ink and right cups - pure water.

The question (problem) is: by a series of such "mixing operations" how much of the water from the left cups can be moved to the right cups ?
It looks like a good problem for high-school competition of a high level, but it is a serious mathematical problem with intriguing connections to many areas of mathematics such as of Majorization Theory, Linear Algebra, Probability Theory, linear dynamical systems, irreversible flows, etc.
The history of this problem started a few years ago when A. Cherny and P. Grigoriev published a paper in one of the most prestigious journals on financial mathematics where they proved the following astonishing statement.

Conditional Expectations as Levels' Averaging

Theorem *Let* (Ω, F, P) be a nonatomic probability space, X, Y two bounded functions with the same dst. Then for any $\varepsilon \geq 0$ there is a sequence of σ-subalgebras $F_1, F_2, ..., F_n \subseteq F$ such that for a sequence of rvs $X_0 = X, X_1 = E(X_0|F_1), X_2 = E(X_1|F_2), ...X_n = E(X_{n-1}|F_n)$, the following inequality holds $||X_n - Y|| \leq \varepsilon$.

In terms of cups it means that it possible to transfer *almost all water* from left to the right if n is sufficiently large. The talk of Cherny at UNCC raised an interesting discussion among A. Gordon, J. Quinn, S. Molchanov and I. Sonin. Almost immediately two other approaches were developed. Later Gordon and Quinn published a paper where ...

The goal of our paper is using our *"hydrostatic"* interpretation to present two simple algorithms describing such transfer and the prove two theorems absent in both papers ChGr and GoQu. First, we describe the structure of all optimal transfers and for all initial position of levels in cups, specified by $2n$ dimensional vector **c**. And second, we give the first exact term of asymptotics of minimal amount that *can not* be transferred to the left

Nonclassical CLT

This term is equal to

$$\frac{1}{\sqrt{\pi n}}. \tag{5}$$

The proof of (5) uses the notion of generating functions and the Central Limit Theorem. This is also pretty remarkable fact that CLT theorem is used for an absolutely deterministic problem.
Let Δ_n be the amount of nontransferable water. It can be shown that

$$\Delta_n = P\{S_1 \geq n\} + ... + P\{S_n \geq n\}, \tag{6}$$

where $S_k = X_1 + ... + X_k$ and rvs X_i are indep. and each having the geometric dst with parameters n and $p = 1/2,$ or in other words S_k has Negative Binomial dst with parameters $k, p = \frac{1}{2}$.
The main goal is to estimate this inequality. The difficulty here lies in the fact that the classical CLT is not working here. We have here normal deviations and large deviations.
Now, why this transfer is optimal ? We replace left cups with girls and the right with boys.

Shall we dance ?

The left subintervals are *girls* and right are *boys*. Each person has a *rank* (level). A rank is equal to the number of "coupons" each person has. The i-th girl has rank $a(i)$ and j-th boy has rank $b(j)$. Every girl can dance only with a boy of lower (equal) rank, i.e. when $a(i) \geq b(j)$.
At each moment of discrete time only one girl will dance with an *available* (for her) boy, i.e. a boy of a lower rank, and after the dance she gives him a coupon and her rank will become $a(i)' = a(i) - 1$ and his rank become $b(j)' = b(j) + 1$. At initial moment an initial distribution of ranks of girls and boys is given and at each moment the state of a system is described by a $2n$ dimensional vector $\mathbf{c} = (\mathbf{a}, \mathbf{b})$, $\mathbf{a} = (a(i), i = 1, 2, ..., n), \mathbf{b}$ $= (b(j), j = 1, 2, ..., n)$. The goal is to organize a *ball*, i.e. select a strategy - a program of dancing, in such way that the maximal number of coupons will be transferred from girls to boys, i.e. the maximal number of dances occur.
An (!) optimal strategy: *At each moment a girl with lowest rank among all open dances with a boy of highest rank available to her.*

Open Problems.

- Finite Nonhomogeneous Markov Chains
- Independence
- Many more

Spasibo !