Estimation of the probability of transition between phases

Jean Claude Dutailly
Paris

21 october 2014

Abstract

The purpose of this paper is to present a general method to estimate the probability law of the transitions between phases of a system. The system must be represented by vectorial variables depending on the time, satisfying some conditions which are usually met. It addresses models in Physics, Economics or Finances.

Quite often, in Economics or physical sciences, one meets systems which, during their evolution, are in distinct types of states, usually called phases. The evolution, which is otherwise continuous, is marked by a discontinuity, a transition between phases. This can be an economic crisis, a rupture in financial markets, an earth-quake, the breakdown of a component in engineering,..The processes which lead to the existence of phases and the passage from one phase to another are various, but usually difficult to model, and the prediction of the transition time, which is usually of paramount importance, cannot be done exactly. So it is common to use probabilist models. The purpose of this paper is to provide a general method to deal with the estimation of such probability laws. It can be implemented for a large class of quantitative models, in economics as well as physical sciences.

In previous papers it has been proven that in quantitative models, using variables X which belong to infinite dimensional vector space - such as scalar functions depending on time - the state of the system can be represented by a vector in a Hilbert space, and as a consequence, this induces many properties of theoretical as well as practical interest. In particular when the variables have time among their parameters and when the process governing the evolution of the system is determinist, it entails that the set $\{X(t), t \in \mathbb{R}\}$ of values taken by the variables can also be endowed with the structure of a Hilbert space F, and that the map which goes from an initial value X(0) to X(t) is an unitary operator. Phases can be characterized by disconnected subsets of the Hilbert space F, and the probability of a phase transition can be legitimately assumed to depend on the distance of a given state to the next phase, distance which has a precise meaning in Hilbert spaces. From there a non parametric method

1

of estimation of the probability law can be deduced, from the knowledge of a past observation of the system.

In a first section the main results of previous papers ("Hilbert spaces in modelling of systems" denoted JCD) are presented. It fixes the general background of the models which are concerned, and notably the conditions under which the results hold.

In the second section we consider the successive steps required to estimate the probability law :

- estimation of the scalar product $\langle \rangle$ in the Hilbert space of values $\{X(t), t \in \mathbb{R}\}$
- the algorithm to define the phases of the system
- computation of the distance of any observation to a phase
- non parametric estimation of the probability law for a phase transition at a given time

1 HILBERT SPACES IN MODELS

1.1 Models

In most scientific theories, quantitative models are used to deal with the analysis of measures, formalization of laws, and computation of predictions. A model comprises a system, the collection of objects which are considered, with their properties represented by a fixed finite number N of quantitative variables $(X_k)_{k=1}^{N}$. Then the state of the system can be defined by the value of these variables, that we will denote collectively X. These variables can be of different kinds, however the results presented in this paper hold when they meet the precise conditions :

Conditions 1:
i) Each variable X_k belongs to an open subset O_k of a separable Fréchet real vector space V_k
ii) At least one of the vector spaces $(V_k)_{k=1}^{N}$ is infinite dimensional

The key points are ;
i) the variables must be vectorial : any linear combination of such X must have a meaning, thus models using qualitative variables (with discrete values) or indices are excluded (but the variables can be smoothed by division by a common constant).
ii) the variables are usually maps (or functions taking values in vector spaces) depending on arguments ξ :
$X_k : \varpi_k \to E_k :: X_k(\xi) = \sum_{i \in I_k} X_{ki} e_{ik}(\xi)$,
the components X_{ki} are fixed scalars, $e_{ik}(\xi)$ are functions taking values in E_k, I_k can be infinite. What is considered is the map X, and not its value X(ξ) for a given value ξ of the argument. This is why they belong to infinite dimensional vector spaces.

The typical example is scalar functions depending on the time t :

$X_k : [0,T] \rightarrow \mathbb{R} :: X_k(t) = \sum_{i \in I_k} X_{ki} e_{ik}(t)$

such as : $X_k(t) = \sum_{n \in \mathbb{Z}} a_n \cos \frac{\varpi nt}{2\pi} + b_n \sin \frac{\varpi nt}{2\pi}$

So the state of the system is defined as the maps X_k : the state is defined all over the domain of the arguments, it is the way the system behave.

The condition i) is quite technical, so let us give some usual examples where the conditions are met :

i) the variables X_k are maps : $X_k : R \rightarrow E$ where R is an open subset of $\mathbb{R}$ and E a finite dimensional vector space, and $\int_R \|X_k(t)\|^2 dt < \infty$ with any norm $\|\|$ on E (then V_k is itself a separable Hilbert space). This addresses :

- all models in analytical Mechanics
- most of the models in fluid mechanics (the variable X can depend on other parameters than t)
- seismic studies (the variables X_k are the waves)
- almost all the models in Economics, representing the evolution of a system : values of bonds, currencies or equities, macroeconomic models, markets studies such as the consumption $X_k(r)$ of products k = 1...N with respect to the income r of a household....

ii) X_k are scalar continuous functions on a compact domain of a topological space

iii) X_k are complex p integrable maps on $\mathbb{R}^n$: $V_k = L^p(\mathbb{R}^n, dx, \mathbb{C})$ with $1 \leq p < \infty$, which addresses most of the models in electromagnetism

A variable X_k and its derivative $\frac{dX_k}{dt}$ are considered as independant variables.

There is no distinction between "internal" and "external" variables : whenever some phenomenon is deemed to have an influence on the state of the system, it is assumed that it can be measured and should be accounted for. If it is external, its value is known beforehand but anyway it can be measured.

A scientific law will usually be expressed by some relation between the variables X_k. The purpose of an experiment, or a test in Economics, is to find or prove a relation between the values of the variables in different specific realisation of the model. Here we stands before the consideration of such laws.

1.2 Hilbert spaces

The first result is the following theorem (see JCD propositions 1 and 2)

Theorem 1 *For any system represented by a model meeting the conditions 1 :*

i) there is a separable, infinite dimensional, Hilbert space H, defined up to isomorphism, such that $S = O_1 \times ... \times O_N$ can be embedded as an open subset $\Omega \subset H$ which contains 0 and a convex subset.

ii) Each state X is associated to a vector ψ of H and there is a linear isometry : $\Upsilon : S \rightarrow \Omega :: \psi = \Upsilon(X)$

So we have the striking result that most models, whatever their field, their objects and hypotheses, have a common mathematical representation. The result stands whatever the domain or the laws which are considered. In particular it does not matter if these laws are determinist or probabilist. For instance if X is a scalar function defined over a period of time [0,T] the function X, if it is square integrable (which is usually the case), belongs to a Hilbert space, whatever the process considered or assumed. The theorem above is just a generalization of this fact.

The theorem seems abstract, but this representation provides powerful tools.

Each variable X_k is itself associated to a Hilbert space H_k and $H = H_1 \times ... \times H_N$

On the real Hilbert space H there are a scalar product denoted $\langle \rangle$, which is a bilinear symmetric form and Hilbertian bases $(\widetilde{\varepsilon}_n)_{n \in \mathbb{N}}$ such that :

$\langle \widetilde{\varepsilon}_n, \widetilde{\varepsilon}_m \rangle = \delta_{mn}, \psi = \sum_{n \in \mathbb{N}} \psi_n \widetilde{\varepsilon}_n, \|\psi\|_F^2 = \langle \psi, \psi \rangle = \sum_{n \in \mathbb{N}} \psi_n^2$

The norm $\|\|_H$ provides the distance between two states : $\|\psi - \psi'\|_H$

The vector space V has also bases $(e_n)_{n \in \mathbb{N}}$, not necessarily orthonormal, such that : $\forall X \in V : X = \sum_{n \in J} X_n e_n$ where only a finite number of components $X_n : n \in J \subset \mathbb{N}$ is non null. But the basis itself has a countably infinite number of vectors.

For any basis $(e_n)_{n \in \mathbb{N}}$ of V there are unique families $(\varepsilon_n)_{n \in \mathbb{N}}, (\phi_n)_{n \in \mathbb{N}}$ of independant vectors of H (which depend on $(e_n)_{n \in \mathbb{N}}$ and are not necessarily Hilbertian) such that :

$\forall m, n \in \mathbb{N} : \langle \phi_n, \varepsilon_m \rangle_H = \delta_{nm}$

and there is a linear bijective map Υ such that X and $\Upsilon(X)$ have the same components respectively in $(e_n)_{n \in \mathbb{N}}, (\varepsilon_n)_{n \in \mathbb{N}}$:

$\forall n \in \mathbb{N} : \varepsilon_n = \Upsilon(e_n)$

$\forall X \in O : X = \sum_{n \in \mathbb{N}} \langle \phi_n, \Upsilon(X) \rangle e_n \rightarrow \Upsilon(X) = \sum_{n \in \mathbb{N}} \langle \phi_n, \Upsilon(X) \rangle \varepsilon_n \in \Omega$

It is clear that in practical experiments variables belonging to infinite dimensional vector spaces cannot be precisely measured (this would require an infinite number of data). So one uses simplified specifications : a variable X is replaced by another Y, which depends on a finite (usually small) number of parameters. For instance a trajectory X(t) is represented by a family of straight lines, or circles, or parabola, depending on a few parameters. More generally an observable is a linear map $\Phi : V \rightarrow V :: Y = \Phi(X)$ (not necessarily continuous) such that its range is a finite dimensional vector subspace of V and $\forall X \in O, \Phi(X)$ is an admissible value, that is $\Phi(O) \subset O$. Notice that this has nothing to do with the precision of the measures : whatever this precision, the scientist gives up the impossible task to estimate X, which is replaced by Y. There are several important results about the values of the observables $\Phi(X)$ with respect to the value of the state X, which are significant mostly in Quantum Mechanics, but they will not be used here.

4

1.3 The evolution of a system

1.3.1 Change of variable in a model

What happens when a given system is represented by different variables ? It depends of course of the nature of the change of variables. As they characterize the system itself, through both the measures which are done and the laws which are tested, a change of variables usually means that one considers a different system (perhaps by focusing the attention to another set of phenomena). But in other cases a change of variable is just a mathematical artefact. For instance this happens when one uses different units. More generally in Physics it happens when the variables are vectorial quantities defined in a frame or are maps depending themselves of coordinates expressed in a frame. A change of frame is mathematically represented by some relation, and one can assume that the use of the new variables should have a distinct impact on the vector ψ representing the state of the system.

We have the general result (JCD proposition 12) :

Theorem 2 *Whenever a system is represented by the variables X belonging to an open subset O of the vector space V, and by the variables X' belonging to an open subset O' of the same vector space V, and there is a continuous, bijective map $U : V \to V$ such that X and X'=U(X) represent the same state of the system then the map U is necessarily linear, there is a unitary, linear, bijective map $\widehat{U} \in \mathcal{L}(H; H)$ such that : $\forall X \in O : \widehat{U}(\Upsilon(X)) = \Upsilon(U(X))$ where Υ is the linear map : $\Upsilon : V \to H$ associated to X.*

The map U is a part of the model itself, and the assumption that X' = U(X) and X represent the same system is of course crucial. In Physics this assumption is usually deduced from more fundamental principles (notably the principle of relativity which says that the laws should not depend on the observer). In Economics there is nothing similar. However a general consequence of this theorem is that the quantities should be represented in the same units : if X' = kX represent the same system, because U is necessarily unitary, then k = 1. If economic data are expressed in different currencies they do not represent the same system.

This result is important, because it helps to find the Hilbert space H : in many cases it belongs to a category linked to the structural definitions of the variables. However for us this result will be implemented in the study of the evolution of a system.

1.3.2 Fundamental theorems

So far we have not been concerned with the laws which are assumed in a model, just by the way it is represented. But when considering the evolution of a system we need to go a bit further. When the time t (it can be any continuous scalar

parameter) is one of the arguments of the maps X, t can be seen with different meanings.

A) t is a parameter used to measure the duration of a phenomenon, usually the time elapsed since some specific event. For instance the fuel left in a tank, the life of an equipment, the elapsed time since the eruption of a volcano,... It is assumed that there are continuous processes at work, with respect to t, whose effects are accounted for by the variables. So obviously X(t) and X'(t) = X'(t+k) with a fixed parameter k do not represent the same phenomenon.

B) t is just a parameter used to identify a state of the system : t gives its temporal location. This is a usual assumption in theoretical physics. The laws do not change with time, all the external factors are accounted for, thus the origin of the time t does not matter. X(t) and X'(t) = X'(t+k) represent the same phenomena.

C) t is still a parameter used to identify each state encountered by the system, but it is not sure that all the phenomena have been accounted for by the model. This is a common case in Economics : that t is expressed by the number of weeks from the beginning of the year, or from the beginning of the last year, should not matter, however usually the week itself matters, because external events may happen which makes that the date itself is significant. Anyway one can assume that since the origin of the time t the processes at work have been continuous, there has been no critical event, and t acquires the meaning of an elapsed time, and we are actually in the case A.

In the case A), B) or C) the purpose of a model is to check or estimate some law : the scientist hopes to find a relation between t and the values X(t) of the variables, which holds whatever the initial values X(0). In the best dream the law would be fully determinist, or at least good enough for the applications. Of course this will not usually be the case, or at least the law will not be fully determinist. It would be impossible to study all the evolution laws that are encountered. So we will proceed the other way around, and look at what happens when the law is determinist. Then one can show that the evolution laws must follow some precise characteristics, from which one can deduce a general method to estimate the transition times. And the discrepancy between what should be if the model was determinist, and what is measured, brings in any case a useful result : it helps to sort out the significant variables. If the measures do not meet the characteristics one can deduce that the variables are either incomplete, or not significant.

Theorem 3 *Whenever in a model :*

i) the variables $(X_k)_{k=1}^N$ are maps : $(X_k)_{k=1}^N :: R \to E$ where R is an open subset of $\mathbb{R}$ (whith origin 0) and E a normed vector space

ii) The map $X = (X_k)_{k=1}^N$ belongs to an open subset O of an infinite dimensional Fréchet space V

iii) $\forall t \in R$ the evaluation map : $\mathcal{E}(t) : O \to E : \mathcal{E}(t) X = X(t)$ is continuous

iv) if for any X,X' the set $\varpi = \{t \in R : X(t) = X'(t)\}$ has a non null Lebesgue measure then $X = X'$

then there is a Hilbert space F, an open $\widetilde{O} \subset F$ such that :

$\forall X \in O \subset V, t \in R : X(0) \in \widetilde{O} \subset F; X(t) \in F$

$\forall t \in R$ the map : $\Theta(t) = \mathcal{E}(t) \circ \mathcal{E}(0)^{-1} \in \mathcal{L}(F;F)$ is unitary

$\forall t \in R$ the map : $\widehat{\mathcal{E}}(t) = \mathcal{E}(t) \circ \mathcal{E}(0)^{-1} \circ \Upsilon^{-1} \in \mathcal{L}(H;F)$ is an isometry

This theorem (JCD 13,14) corresponds to the cases A) and C). The condition iv) is the most important : it states that the laws governing the evolution of the system are semi-determinist. They provide a solution, unique up to a non null set of values of t. One cannot have, for the same initial conditions, two significantly different solutions.

As a consequence :

- for any $u \in \widetilde{O}$ then $\forall t : X(t) = \Theta(t)(u)$ is well defined and X(t) depends continuously on u : one says that the problem is well posed

- for any value $u \in F, t_0 > 0 \in R$ there is a unique map $X \in V$ such that $X(t_0) = u$ defined by $X = \Upsilon \circ \mathcal{E}(t_0) \circ \mathcal{E}(0)^{-1} \circ \Upsilon^{-1}(u)$: there is a unique map X which goes to a given point. More precisely if there were two such maps they would take the same values almost everywhere in R.

Notice that the theorem (as well as the previous ones) does not require any continuity conditions on the maps X, and there is no probability involved. The conditions i) and ii) are similar to the conditions 1, but E must be a normed vector space (possibly infinite dimensional). So X(t) can itself be a map, depending of other arguments. The condition iii) is rather technical, and is usually met.

The cases of type B can be specified by the condition that the variables $X'(t) = X(t+\theta)$ and $X(t)$ represent the same state of the system, and we have the stronger result (JCD 15,16) :

Theorem 4 *Whenever in a model :*

i) the variables $(X_k)_{k=1}^{N}$ are maps : $(X_k)_{k=1}^{N} :: R \to E$ where E is a normed vector space

ii) The map $X = (X_k)_{k=1}^{N}$ belongs to an open subset O of an infinite dimensional Fréchet space V

iii) $\forall t \in \mathbb{R}$ the evaluation map : $\mathcal{E}(t) : O \to E : \mathcal{E}(t)X = X(t)$ is continuous

iiv) for any fixed $\theta \in R$, the variables $X'_k(t) = X_k(t+\theta)$ and $X_k(t)$ represent the same state of the system

then :

there is a Hilbert space F, an open $\widetilde{O} \subset F$, a continuous anti-hermitian map $\widetilde{S} \in \mathcal{L}(F;F)$ such that :

$\forall X \in O \subset V : X(0) \in \widetilde{O} \subset F$

$\forall t : X(t) = \left(\exp t\widetilde{S}\right)(X(0)) \in F$

The maps X are smooth and $\frac{d}{ds}X(s)|_{s=t} = \widetilde{S}X(t)$
There is a continuous map $S \in \mathcal{L}(V;V)$ such that :
$\Theta(t) = \mathcal{E}(t) \circ \mathcal{E}(0)^{-1} = \exp tS$
$\forall t \in \mathbb{R} : X(t) = (\exp tS \circ X)(0) = \left(\sum_{n=0}^{\infty} \frac{t^n}{n!} S^n X\right)(0)$
and the operator $\widehat{S} = \Upsilon \circ S \circ \Upsilon^{-1}$ associated to S is anti-hermitian

Actually it is easy to see that this is a special case of the previous theorem, and Θ is then defined by the exponential of a unique operator. So, not only the model is determinist, we have a generalized exponential law. This is intuitive : if the system is in a steady environment, it should have a simple evolution, with some constant law. This law is given by a single anti-hermitian map $\widetilde{S}$ which is directly related to the derivative $\frac{d}{ds}X(s)$. Notice that, even if X was not assumed to be continuous, smoothness is a necessary result. This is the starting point for the Schrodinger equation of Quantum Mechanics.

1.3.3   Phases

So, when the model is determinist, the evolution laws shall meet some precise conditions, in particular in the case B), whatever the processes at work. There is a large class of problems where the maps X belong to the same family but the states X(t), for some periods, take significantly different values in the same vector space E : the system meets a phase transition. The conditions in which these transitions happen are of special interest. Common cases in Physics are change of phases for solid or liquid bodies, the desintegration of a particle, an earthquake,..., in Economics a crisis or a recession, in Finances a flip in the markets,...The questions which arise are then : what are the conditions, about the initial conditions or the maps X, for the occurence of such an event ? Can we forecast the time at which such event takes place ?

The states of the system are represented by vectors of E, and so it is legitimate to characterize the phases as connected components of the set E. Connectedness has a precise definition in Mathematics, and it can be applied in E. Roughly two different phases are sets which have no common point.

With this definition, as a first consequence, the map X cannot be continuous (if it were, then the set of values $\{X(t), t \in R\}$ would also be connected) and one cannot be in the case B : phases transitions can exist only if there is a discontinuity somewhere, either in the laws governing the evolution or if there is a change in the environment of the system. This is in itself an important result : one cannot expect to represent discontinuous processes with models of the type B). One needs either additional variables, for instance accounting for external events, or assume that the laws are themselves discontinuous, which is always a difficult issue, at least at the theoretical level.

Totally discontinuous maps exist, but they are strange mathematical objects. Usually discontinuities happen at isolated points : the existence of a singularity

is what makes interesting a change of phase. If the transition points are isolated, then there are only at most a countable number of them. A given map X is then continuous (with respect to t) except in a set of points t_p, which is finite over any finite period, and we have a series of phases separated by transitions occurring at precise times. So one can safely say that any scientific sensible model depicts either continuous evolutions or a finite number of phases, each one corresponding to a continuous evolution, separated by precise instantaneous transitions.

From this it is legitimate to estimate the law of evolution X(t) by a family of maps $X_n(t) = \left(\exp t \widetilde{S}_n \right) (X_n(0))$ with different, but constant, values $\widetilde{S}_n$ during each phase. Each operator $\exp t \widetilde{S}_n$ is unitary, but this procedure is not easy because they do not constitute either a vector space, or a subgroup. So to study the transitions we will use another way.

Usually it is not possible to predict exactly a change of phase, and transitions are formalized by a probability law. A sensible assumption is that the probability for a change of phase depends on the proximity of the state of the system from each phase. If we are in the conditions of the theorem 3 it is possible to address practically the problem.

Let us consider two phases, characterized by disjoint subsets E_1, E_2. Their choice is somewhat arbitrary, and anyway would be adjusted from previous data. So we have two disjoint subsets F_1, F_2 of the Hilbert space F. If F_1, F_2 are *closed convex subsets* of F the distance of any point x of F to one of the set F_i is defined by the projection $\pi_i : F \to E_i$: there is a unique $y = \pi_i(x) \in E_i$ such that $\|x - y\|_F$ is minimum. The map π_i is continuous, $\pi_i^2 = \pi_i$ and $\pi_i(x) = x$ when $x \in F_i$. So when we are in the phase F_1 we can relate the probability of a transition 1 -> 2 to $\|X(t) - \pi_2(X(t))\|_F$. And more generally the probability of any transition can be related to the quantity $r(t) = \|X(t) - \pi_1(X(t))\|_F + \|X(t) - \pi_2(X(t))\|_F$.

The result holds if F_1, F_2 are closed *vector subspaces* of F such that $F_1 \cap F_2 = \{0\}$. Then

$X(t) = \pi_1(X(t)) + \pi_2(X(t))$

and $\|X(t)\|^2 = \|\pi_1(X(t))\|^2 + \|\pi_2(X(t))\|^2$

$\frac{\|\pi_1(X(t))\|^2}{\|X(t)\|^2}$ can be interpreted as the probability that the system at t is in the phase F_1.

We have now the material to proceed to the core of this paper : if we assume that the probability of a transition at a time t is a function f of r(t) then there is a general method to estimate f. The key step is to compute the function $d : F \times F \to \mathbb{R}$ which gives the distance between two points of F, using the metric induced by the scalar product in the Hilbert space F.

2 ESTIMATION OF THE PROBABILITY OF A TRANSITION

2.1 Distance between states

The map $X : [0, T] \to E : X(t)$ takes its value in the normed space E, which can be infinite dimensional. Due to the evolution laws X(t) belongs to a vector subspace E_0 of E, which is completed to get the Hilbert space F. So F itself is not necessarily included in E, but any realisation X(t) can be seen as belonging to E and F. F is endowed with a scalar product, denoted $\langle \rangle$ usually unknown. Even if E is a normed vector space, and if it is possible to deduce (by polarization) a scalar product from its norm (which is not always possible), there is no guarantee that it would be the same as $\langle \rangle$. So the goal is to estimate this scalar product, from where the function d is computed :

$d : F \times F \to \mathbb{R} :: d(X, X') = \sqrt{\langle X - X', X - X' \rangle}$

from a set of observations on a past period.

A key point in the demonstration of the theorem 3 is that, because any point X(t) can be identified with a map X, the scalar product on F can be defined by :

$\langle X(t), X'(t') \rangle_F = \langle \Upsilon X, \Upsilon X' \rangle_H$

so that for any observations at t_p, t_q :

$\forall p, q \leq N : \langle X(t_p), X(t_q) \rangle_F = \langle \Upsilon X, \Upsilon X \rangle_H = Ct = K$

for a given state X of the system.

The theorem 3 holds when E is an infinite dimensional vector space, which happens if t is not the only parameter, and other parameters, denoted ξ, are considered. But then, from a finite set of observations at given times $(t_p)_{p=1}^N$ it is usually impossible to have an efficient estimate of the maps $X(t, \xi)$: the parameters $(\xi_p)_{p=1}^N$ take fixed values which are not sufficient to approximate the maps $X(t, .)$. So we will limite ourselves to the case, which should be the most usual, when the vector space E (and then F which is the completion of one of its vector subspace) has the finite dimension m. Then the general form of the maps X is : $X(t) = \sum_{p=1}^m X_p(t) e_p$ where $(e_p)_{n=1}^m$ are fixed vectors and X_p are scalar functions, depending on t only. These functions belong themselves to a infinite dimensional vector space : $X_p(t) = \sum_{i \in I} x_{pi} E_i(t)$ where $(E_i)_{i \in I}$ defines a basis of V.

It is assumed that one has a set $\{X(t_n)\}_{n=1}^N$ of observations, on a past period, for the system.

The realisations X(t) always belong to F and E, so that we will take E as a proxy for F itself. Then the scalar product $\langle \rangle_F$ is represented in any basis of E by a symmetric, definite positive, m×m matrix B, and, in matricial notations :

$\forall p, q : \langle X(t_p), X(t_q) \rangle_F = [X(t_p)]^t [B] [X(t_q)] = K$

Which reads:

$[X]^t [B] [X] = K [U]$

where $[X]$ is the m×N matrix of the measures, and [U] is the N×N matrix comprised of 1. K is an unknown quantity, but it is clear that B, K are defined up to a scalar positive constant so, for the time being, let us keep K as a parameter.

The vector space of square matrices is normed, with the norm :

$$\|M\|^2 = Tr\,[M]^t\,[M] = \sum_{ij}\left([M]_j^i\right)^2$$

We choose as estimation of B the matrix $\widehat{B}$ such that the distance between $K\,[U]$ and $[X]^t\,[B]\,[X]$ is minimum :

$$\widehat{B} : \min_B \left\|[X]^t\,[B]\,[X] - K\,[U]\right\|^2$$

Because $[X]^t\,[B]\,[X]\,, K\,[U]$ are symmetric it reads :

$$\left\|[X]^t\,[B]\,[X] - K\,[U]\right\|^2 =$$

$$= Tr\left([X]^t\,[B]\,[X] - K\,[U]\right)^2$$

$$= Tr\left([X]^t\,[B]\,[X]\,[X]^t\,[B]\,[X]\right) - KTr\left([X]^t\,[B]\,[X]\,[U]\right)$$

$$-KTr\left([U]\,[X]^t\,[B]\,[X]\right) + K^2 Tr\,[U]^2$$

$$= Tr\left([X]\,[X]^t\,[B]\,[X]\,[X]^t\,[B]\right) - KTr\left([X]\,[U]\,[X]^t\,[B]\right)$$

$$-KTr\left([X]\,[U]\,[X]^t\,[B]\right) + K^2 N^2$$

$$= Tr\left([P]\,[B]\,[P]\,[B]\right) - 2KTr\left([Q]\,[B]\right) + K^2 N^2$$

where $[P] = [X]\,[X]^t = [P]^t\,, [Q] = [X]\,[U]\,[X]^t = [Q]^t$ are m×m known symmetric matrices.

B is a symmetric, definite positive, matrix, which reads : $B = A^t A$ where A is a m×m matrix. $[A]\,[X\,(t_q)]$ is the matrix column of the components of $X\,(t_q)$ in an orthonormal basis :

$$[X\,(t_p)]^t\,[B]\,[X\,(t_q)] = [X\,(t_p)]^t\,[A]^t\,[A]\,[X\,(t_q)] = K$$

In components :

$$\left\|[X]^t\,[B]\,[X] - \widehat{K}\,[U]\right\|^2$$

$$= \sum_{ijklpq=1}^m [P]_j^i\,[A]_j^k\,[A]_l^k\,[P]_p^l\,[A]_p^q\,[A]_i^q - 2K\sum_{p,q=1}^N \sum_{i,j,k=1}^m X^j\,(t_q)\,[A]_j^k\,[A]_i^k\,X^i\,(t_p)$$

The derivative with respect to $[A]_\beta^\alpha$ is:

$$\sum_{ijklpq=1}^m [P]_\beta^i\,[A]_l^\alpha\,[P]_p^l\,[A]_p^q\,[A]_i^q + [P]_j^i\,[A]_j^\alpha\,[P]_p^\beta\,[A]_p^q\,[A]_i^q$$

$$+ [P]_j^i\,[A]_j^k\,[A]_l^k\,[P]_\beta^l\,[A]_i^\alpha + [P]_j^\beta\,[A]_j^k\,[A]_l^k\,[P]_p^l\,[A]_p^\alpha$$

$$-2K\sum_{p,q=1}^N \sum_{i,j,k=1}^m \left(X^\beta\,(t_q)\,[A]_i^\alpha\,X^i\,(t_p) + X^j\,(t_q)\,[A]_{ij}^\alpha\,X^\beta\,(t_p)\right)$$

$$= \sum_{ijklpq=1}^m [A]_l^\alpha\,[P]_p^l\,[A]_p^q\,[A]_i^q\,[P]_\beta^i + [A]_j^\alpha\,[P]_j^i\,[A]_p^q\,[A]_i^q\,[P]_p^\beta$$

$$+ [A]_i^\alpha\,[P]_j^i\,[A]_j^k\,[A]_l^k\,[P]_\beta^l + [A]_i^\alpha\,[P]_p^l\,[A]_l^k\,[A]_j^k\,[P]_j^\beta$$

$$-2K\sum_{p,q=1}^N \sum_{i,j,k=1}^m \left([A]_i^\alpha\,X^i\,(t_p)\,X^\beta\,(t_q) + [A]_j^\alpha\,X^j\,(t_q)\,X^\beta\,(t_p)\right)$$

$$= 4\left([A]\,[P]\,[A]^t\,[A]\,[P]\right)_\beta^\alpha - 4K\sum_{p,q=1}^N [A]\,[X\,(t_p)]\,[X\,(t_q)]^t$$

The scalar product must be non degenerate, so $\det B \neq 0 \Rightarrow \det A \neq 0$ and by multiplication with A^{-1} the condition for the minimum reads in matricial notation :

$[P] [A]^t [A] [P] = K \sum_{p,q=1}^{N} [X(t_p)] [X(t_q)]^t$

$[P] [B] [P] = K \sum_{p,q=1}^{N} [X(t_p)] [X(t_q)]^t$

For any $\left[\widehat{B}\right]$ meeting this condition :

$\left\| [X]^t \left[\widehat{B}\right] [X] - K [U] \right\|^2$

$= Tr\left([P] \left[\widehat{B}\right] [P] \left[\widehat{B}\right] \right) - 2KTr\left([Q] \left[\widehat{B}\right] \right) + K^2 N^2$

$= KTr\left(\sum_{p,q=1}^{N} [X(t_p)] [X(t_q)]^t \left[\widehat{B}\right] \right) - 2KTr\left([Q] \left[\widehat{B}\right] \right) + K^2 N^2$

$Tr\left(\sum_{p,q=1}^{N} [X(t_p)] [X(t_q)]^t \left[\widehat{B}\right] \right)$

$= \sum_{p,q=1}^{N} Tr\left([X(t_p)] [X(t_q)]^t \left[\widehat{B}\right] \right)$

$= \sum_{p,q=1}^{N} \left(\sum_{i,j=1}^{m} [X(t_p)]^i [X(t_q)]^j \left[\widehat{B}\right]_i^j \right)$

$= \sum_{p,q=1}^{N} \left(\sum_{i,j=1}^{m} [X(t_p)]^i \left[\widehat{B}\right]_j^i [X(t_q)]^j \right) = \sum_{p,q=1}^{N} [X(t_p)]^t \left[\widehat{B}\right] [X(t_q)]$

$[Q]_j^i = \left([X] [U] [X]^t \right)_j^i = \sum_{p,q=1}^{N} [X]_p^i [U]_q^p [X]_j^q = \sum_{p,q=1}^{N} X^i(t_p) X^j(t_q)$

$Tr\left([Q] \left[\widehat{B}\right] \right) = \sum_{i,j=1}^{m} \sum_{p,q=1}^{N} X^i(t_p) X^j(t_q) \left[\widehat{B}\right]_i^j$

$= \sum_{p,q=1}^{N} \sum_{i,j=1}^{m} X^j(t_q) \left[\widehat{B}\right]_i^j X^i(t_p) = \sum_{p,q=1}^{N} [X(t_q)]^t \left[\widehat{B}\right] [X(t_p)]$

$\left\| [X]^t \left[\widehat{B}\right] [X] - K [U] \right\|^2$

$= K \sum_{p,q=1}^{N} [X(t_p)]^t \left[\widehat{B}\right] [X(t_q)] - 2K \sum_{p,q=1}^{N} [X(t_q)]^t \left[\widehat{B}\right] [X(t_p)] + N^2$

$= N^2 - K \sum_{p,q=1}^{N} [X(t_q)]^t \left[\widehat{B}\right] [X(t_p)]$

If we have, as expected, $[X(t_q)]^t \left[\widehat{B}\right] [X(t_p)] = K$

$\left\| [X]^t \left[\widehat{B}\right] [X] - \widehat{K} [U] \right\|^2 = N^2 - KN^2$

The minimum is 0 for : $K = 1$

Thus the estimate of B is, if $[P] = [X] [X]^t$ is invertible :

$$\left[\widehat{B}\right] = \left([X] [X]^t \right)^{-1} \left(\sum_{p,q=1}^{N} [X(t_p)] [X(t_q)]^t \right) \left([X] [X]^t \right)^{-1} \tag{1}$$

This is a symmetric matrix. Moreover

$\left([X] [X]^t \right)_j^i = \sum_{p,q=1}^{N} [X(t_p)]^i [X_q]^j$

thus $\left[\widehat{B}\right]$ does not depend on the time ordering.

2.2 Defining the phases

Usually the periods characteristic of the phases are defined, up to one or two fringe observations. A principal component analysis of the set of points given

by X may help to define the subsets corresponding to the different phases (see Kendall III chap.44). But here we have two tools to proceed to a clustering of the data : the distance given by B, and the time ordering. Indeed we expect from the phases that they provide data that are both close in distance, and successive in time. The method suggested below, sequential clustering, is one among others that can be considered.

A cluster C is a set of successive observations. The clusters are indexed by two scalars r and k, and C(r,k) is a table with the first θ_1 and the last θ_2 of the ordered times of the observations in the cluster : all observations such that $\theta_1 \leq t_p \leq \theta_2$ belong to C(r,k), and $\theta_1 \leq \theta_2$. Initially r = 1 and $C(1,n) = [t_n, t_n]$ with n = 1 ... N.

We consider the N×N matrix : $D = [d\left(X\left(t_p\right), X\left(t_q\right)\right)]_{p,q=1}^{N}$ and the set $\{\delta_p\}_{p=1}^{P}$ of ordered values of $d\left(X\left(t_p\right), X\left(t_q\right)\right)$:

$\delta_1 = \min_{p,q=1..N} d\left(X\left(t_p\right), X\left(t_q\right)\right)$

$\delta_1 \leq \delta_2 \leq ... \leq \delta_P$

The distance between two clusters C(r,i), C(r,j) is defined as :

$\Delta\left(C\left(r,i\right), C\left(r,j\right)\right) = \min_{t_p \in C(r,i), t_q \in C(r,j)} d\left(X\left(t_p\right), X\left(t_q\right)\right)$

One proceeds by steps denoted r.

Fix $\delta = \delta_1$

Review $X\left(t_n\right)$ from n = 1 to n = N

Merge t_2 in the cluster $C\left(1,1\right) = [t_1, t_2]$ if $d\left(X\left(t_1\right), X\left(t_2\right)\right) \leq \delta$, and put $C\left(1,2\right) = C\left(1,1\right)$

Merge t_3 in the cluster $C\left(1,2\right)$ if $\Delta\left(C\left(1,3\right), C\left(1,2\right)\right) \leq \delta$, and put $C\left(1,3\right) = C\left(1,2\right)$

...

Then on the next step put $\delta = \delta_2$, reinitiate $C\left(2,i\right) = C\left(1,i\right)$ and proceeds similalrly.

Thus, in a finite number of steps, all the observations are clustered.

One can choose the clustering provided at any step r : the clusters comprise successive data, which are such that the distance between them is at most δ_r. Of course it is possible that some points are closer but do not belong to the same cluster, this case should occur normally if the same phase appears more than once. And some points may stay isolated : they should be either discarded or incorporated in one of the existing clusters. The phases are then chosen from clusters which have values which are close. With this procedure one can keep a control of the clustering, and check that the initial assumptions - that there exist phases - is grounded.

A phase is a convex, closed, subset of points in F. So if we have a family of points $\{x_p\}_{p=1}^{n}$ it defines the convex set as all the points :

$x = \sum_{p=1}^{n} \xi_p x_p, 0 \leq \xi_p \leq 1$

In the phases defined by the clustering there can be interior points (which are a linear combination of the others) and they can be removed. However if

the dimension m of E is large this is not easy, and anyway to keep them entails only a waste of computation time.

2.3 Computing the distance to a phase

From general theorems about Hilbert space, any point y as a unique projection $\pi(y)$ on a closed convex subset C, defined by :

$d(y, \pi(y)) = \min_{x \in C} d(y, x)$

and the distance from y to C is defined as :

$\Delta(y, C) = d(y, \pi(y))$

It is null if $y \in C$.

So for any point y its projection onto a phase defined by the family $\{x_p\}_{p=1}^n$ is computed by looking for the minimum of the function :

$d\left(y, \sum_{p=1}^n \xi_p x_p\right)$ for $0 \leq \xi_p \leq 1$

We know that there is a unique solution.

$d\left(y, \sum_{p=1}^n \xi_p x_p\right) = \sqrt{\left[y - \sum_{p=1}^n \xi_p x_p\right]^t B \left[y - \sum_{p=1}^n \xi_p x_p\right]}$

$= \sqrt{\sum_{i,j=1}^m \left[y^i - \sum_{p=1}^n \xi_p x_p^i\right] B_j^i \left[y^j - \sum_{p=1}^n \xi_p x_p^j\right]}$

The derivative of d^2 with respect to ξ_p is :

$\sum_{i,j=1}^m \left[-x_p^i\right] B_j^i \left[y^j - \sum_{q=1}^n \xi_q x_q^j\right] + \left[y^i - \sum_{q=1}^n \xi_q x_q^i\right] B_j^i \left[x_p^j\right]$

$= -[x_p]^t B \left[y - \sum_{q=1}^n \xi_q x_q\right] - \left[y - \sum_{q=1}^n \xi_q x_q\right]^t B [x_p]$

which provides the set of n linear equations :

$\sum_{q=1}^n \xi_q [x_p]^t B [x_q] = [x_p]^t B [y]$

and the minimum is :

$\left[y - \sum_{p=1}^n \xi_p x_p\right]^t B \left[y - \sum_{p=1}^n \xi_p x_p\right]$

$= [y]^t B [y] - \sum_{p=1}^n [\xi_p x_p]^t B [y] - \sum_{p=1}^n [y]^t B [\xi_p x_p] + \sum_{p,q=1}^n \xi_p \xi_q [x_p]^t B [x_q]$

$= [y]^t B [y] - 2\sum_{p,q=1}^n \xi_p \xi_q [x_p]^t B [x_q] + \sum_{p,q=1}^n \xi_p \xi_q [x_p]^t B [x_q]$

$(d(y, \pi(y)))^2 = [y]^t B [y] - \sum_{p,q=1}^n \xi_p \xi_q [x_p]^t B [x_q]$

If the projection is an interior point then $0 < \xi_p < 1, p = 1..n$. The probability that the projection is on some edge is low, however if there is no interior solution one must check with the extremal values $\xi_p = 0, 1$ as in any linear programming problem. The computation must be done for each phase. If y belongs to the phase then $d(y, \pi(y)) = 0$.

2.4 Estimation of the probability of a transition

Let us consider a system with two phases, defined by the convex subsets C_1, C_2.

The assumptions are the following :

i) At any given time the state X(t) of the system is either in C_1 or in C_2

ii) The probability that a transition occurs at t is a function of :

$r(t) = \Delta(X(t), C_1) + \Delta(X(t), C_2)$

Because of i) one of the two quantities $\Delta\left(X\left(t\right),C_1\right),\Delta\left(X\left(t\right),C_2\right)$ is null, so r(t) is the distance from X(t) to the other phase.

The density of probability of a transition at any given time t is :

Pr(transition at t $\mid r\left(t\right)=r$) $=f\left(r\right)$ so it does not depend on t, only on the value of r(t)

We assume as above that the system has been followed on a period of time [0,T] during which some transitions have been observed. This is of course a critical condition, and the quality of the estimation depends of the number of observed transitions. As above there are N observations, and we assume that they ere evenly spaced.

For each observation at t_p one can compute $r\left(t_p\right)$ as above, and the number $\mu\left(t_p\right)$ (it is 0 or 1) of transitions which have occured in the following period $\left[t_p,t_{p+1}\right]$. The total number of transitions over [0,T] is μ_T.

The probability that a transition occurs during the next period is then estimated by $\pi\left(r\left(t_p\right)\right)=\mu\left(t_p\right)/\mu_T$. The graph of the points $\left(r\left(t_p\right),\pi\left(r\left(t_p\right)\right)\right)$ gives a curve, from which the density f can be estimated. Because the number of transition is usually low, it is more accurate to estimate the probability of a transition if $r\left(t\right)\leq r$: take all the observations such that $r\left(t_p\right)\leq r$ for different values of r, and add the corresponding $\pi\left(r\left(t_p\right)\right)$.

The curve $\left(r,\pi\left(r\right)\right)$ should be decreasing (even if this condition is not used in the computation above). Anyway the simple view of the curve (t, r(t)) should show the discontinuities happening at the transition points.

2.5 Comments and conclusion

1. The computation and estimation above are based on the theorem 3 and several assumptions, whose validity can be checked from the data.

i) If the clustering does not show definite phases, one can legitimately assume that the model is not accurate. The variables used are not pertinent to determine the states of the system, at least in what is considered as significant periods.

ii) If the clustering provides significant phases, but the graph of the curve (t, r(t)) shows no discontinuities between the phases, then the assumption about the probability of transition does not hold. The prospect of finding a predictive model with these variables is not good : usually the discontinuities come from external events, which are not accounted for.

iii) If the clustering and the curve (t, r(t)) shows clear discontinuities then the estimate of f(r) is a sensible choice. But of course its quality depends greatly of the number of transitions which have been observed, and of their similarities.

So, overall the method is valid as long as one has some continuity in the processes at work, and that their impact is well represented by the variables X. But it helps to sort out the potential variables.

2. This raises an issue : is it ever possible to be in a situation where all these conditions are met ? One can extend a model, add other variables, refine

their definitions so that they account more accurately for all kinds of initial conditions. But one can expect that, in all practical cases, there will be variables which have been omitted. Usually their impact is summarized by probabilist laws. The method exposed here holds in this case : the unknown variables Y are such that the set (X,Y) gives a determinist law. The evaluation of B should account for Y, for which a distribution law is assumed. As a consequence $\widehat{B}$ follows a probability law, depending on Y and the usual stattistical methods can be implemented to precise the characteristics of the estimator.

3. Using the results exposed in the section 1 on can shows that interacting systems take specific states : this is not only their measures which show a definite pattern, the interactions constrain the systems to stay all together in some coherent states. The interactions do not bring chaos, they bring order. And this is usually the processes which are at the root of the existence of phases. Because these processes occur at an elementary level, and imply a great number of interacting systems, it is difficult or impossible to formalize explicitely what happens at a macroscopic level. And this motivates the search for non fully determinist laws, involving only measurable quantities at the observable level.

3   REFERENCES

J.C.Dutailly *Hilbert spaces in modelling of systems* https://hal.archives-ouvertes.fr/hal-00974251 - (5 apr 2014)

M.Kendall *The advanced theory of statistics* C.Griffin (1983)