

Probabilistic Information Transfer
by Olga Zeydina and Bernard Beauzamy

Readers' comments (in English and in French)

De Thibault Quiviger, 25 juin 2013 :

Le livre de Olga Zeydina et Bernard Beauzamy sur la construction de l'hypersurface probabiliste est précieux à plusieurs titres :

Il réconcilie les ingénieurs avec les probabilités : les probabilités sont souvent délaissées par nous autres ingénieurs, qui préférons des réponses quantitatives précises (et fausses) plutôt que la responsabilité du choix laissé par une approche probabiliste. Les ingénieurs utilisent trop souvent des techniques d'optimisation pointues et instables tout en s'appuyant sur des données imprécises.

Il souligne l'arbitraire de nos approches habituelles ; il est tellement pratique de supposer une variable gaussienne ou exponentiellement décroissante alors que rien ne devrait nous pousser à cette facilité coupable. Certes, nous connaissons la fragilité de nos modèles mais finalement, bombardés experts, dans l'industrie, bien peu pourront nous contredire. La vraie élégance intellectuelle consiste à ne tenir compte que des données et ne pas tordre la réalité pour la conformer aux canons simplistes de nos études.

Il montre que même dans des situations où les données sont rares, il est possible de déterminer des réponses probabilistes à partir des données présentes et uniquement de ces dernières, sans faire absolument aucune autre hypothèse que celle de la maximisation de l'entropie ou autrement dit, notre méconnaissance du système. C'est notre méconnaissance du système, et elle seule, qui doit nous guider.

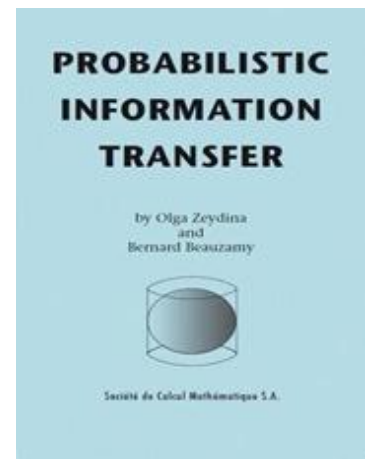
Enfin, cette approche montre aussi les limites de la reconstruction de l'information à partir de l'information existante. Il arrive parfois que cette information brute soit insuffisante et exige d'incorporer des lois physiques, économiques, mathématiques autres que celle de la maximisation de l'entropie, pour pouvoir résoudre le problème posé. Au moins, ce sera alors en connaissance de cause et avec toutes les précautions d'usage, connaissant maintenant l'arbitraire de la technique.

Thibault Quiviger
Maître de conférences associé à l'UTC
Président d'Orcades Commodities
Consultant en modélisation des prix des matières premières

From Thibaut Quiviger, June 25th 2013

The book by Olga Zeydina and Bernard Beauzamy about the Probabilistic Information Transfer is unique and precious in many ways:

It makes engineers and probabilities friends again : as engineers, we often discard probabilities, preferring sharp (and wrong) quantitative answers instead of the



responsibility of making a decision based on probabilities. Too often engineers resort to advanced and unstable optimization techniques while basing them on scarce and blurred data.

It emphasizes how arbitrary are our usual resolution techniques ; it is so easy to pose that a given variable is Gaussian or exponential whereas nothing should let us go this easy and guilty way. Of course we know how fragile are our models but, as so called experts, few will dare contradict us. The true intellectual elegance lies in considering the raw data only and not in distorting the reality so that it fits the usual models we learnt to solve at the university.

It shows that even when data are scarce, it is possible to compute probabilistic answers from the data available and only from them, without any external assumption but the entropy maximization. That is to say, only our maximum ignorance should lead us in the search for numbers.

Last but not least, the EPH shows also the limits of rebuilding information out of the data available. It may happen that the raw information available is not enough. Besides the entropy, new laws must be then incorporated from different fields (physics, mathematics,...) so that we get satisfactory answers. But now, we know how biased and arbitrary this can be and we will do it every time more carefully.

Thibault QUIVIGER
CEO Orcades COMmodities
System modeling and Commodity pricing expert

http://www.cnrs.fr/dire/actualites/2013/juillet/ouvrage_280613.htm

"Probabilistic Information Transfer", nouvel ouvrage sur les mathématiques du réel

Olga Zeydina et Bernard Beauzamy nous proposent un sixième volet de la collection de "Les mathématiques du réel", destinée aux chercheurs et aux ingénieurs. L'ouvrage "Probabilistic Information Transfer" édité par la Société de calcul de mathématique en avril 2013, montre comment "propager" l'information à partir d'une situation existante vers une situation de besoin. L'information ainsi propagée se détériore avec la distance, comme un champ gravitationnel décroît avec la distance.

Dans les situations de la vie réelle, on n'a jamais l'information détaillée que l'on souhaiterait avoir : elle est souvent incomplète ou corrompue. Mais, à l'inverse, on dispose généralement de données éparses. Il y a donc un besoin naturel : utiliser ce que l'on possède pour reconstituer ce qui manque. Ceci doit être fait en respectant deux règles :

- Il ne faut pas ajouter d'information artificielle, comme des hypothèses de modèle (déclarer que telle croissance est linéaire, que telle loi est gaussienne, etc.) ;
- Le résultat doit être de nature probabiliste : nous ne voulons pas une valeur précise (qui n'a pas de sens), mais une loi de probabilité, avec des intervalles de confiance, comme le réclament les autorités de sûreté.

Le livre est organisé en trois parties : les constructions de base, (accessibles sans aucune connaissance particulière en probabilités), les applications en situation réelle et la théorie.

Cette théorie mathématique nouvelle, entièrement développée par la société de calcul mathématique, répond à un besoin initialement exprimé par un industriel, Framatome-ANP, en 2003. Les applications déjà recensées sont nombreuses : classification d'objets industriels (Air Liquide), évaluation de pollutions (Total), estimation de la qualité de l'eau dans des rivières (Agence Européenne de l'Environnement), application à la sécurité des réacteurs (Institut de Radioprotection et de Sûreté Nucléaire), etc.

Pour en savoir plus: http://scmsa.eu/archives/SCM_PIT_order.htm

Contact : Bernard Beauzamy / Société de Calcul Mathématique S.A / T 01 42 89 08 91 / bernard.beauzamy@scmsa.com

From Giovanni Bruna, Scientific Director, Institut de Radioprotection et de Sûreté Nucléaire, 30/07/2013

Probabilistic Information Transfer, cosigned by Olga Zeydina & Bernard Beauzamy and printed by Normandie Roto, Alençon, is a 208 page book, the sixth in the SCM's Collection on Real Life Mathematics. It is the second book written jointly by the authors after the very attractive *Méthodes probabilistes pour la reconstruction de données manquantes* (Probabilistic methods for reconstruction of missing data), which widely relies upon the relevant Olga Zeydina's thesis work.

The book addresses the crucial problem of the reliability of information: it wonders upon its very nature and use for scientific and industrial purposes.

In everyday life, information is generally abundant and more and more freely available. But is it pertinent? Is it sufficiently comprehensive and reliable for scientific and technical uses?

The authors answer that it is generally not the case: available data are often incomplete and somewhat sparse, sometimes corrupted and non-reliable, even false, they claim. Moreover, even if over-abundant, contrary to the common sense, data could not fully match the needs of end-users.

Thus, before any practical use, data have to be critically analyzed and investigated: sometimes they have to be selected and filtered, sometimes complemented, anytime they must be harmonized and fitted to the targeted objectives. That has to be done strictly respecting some elementary rules and constraints aimed at avoiding addition of spurious information and its corruption through the assumptions and the treatment process.

As far as the process of complementing data is concerned, the authors claim that at least two main working-rules have to be adopted:

- Never binding the problem through the assumptions built-in in the models used to handle them (either implicit or explicit ones), e.g., through assuming or postulating symmetry conditions, linear behavior, Gaussian laws, and so on;

– Searching for probability laws with confidence intervals instead of absolute values that could turn-out nonsense, in practice.

The book, which widely relies upon previous SCM's works, the rich content of which has been very opportunely critically reinvested and systematized, describes the authors' original methodology to address the crucial problem of complementing available information to obtain a comprehensive-enough one for practical uses.

The method, originally developed for industrial use in the nuclear field in the framework of a 2003 contract with Framatome-ANP, has been since then getting applications in different fields of endeavor, such as the classification of industrial objects (Air Liquide), the evaluation of an harbor site pollution (Total), the quantification of water purity level in the rivers (European agency for Environment), the appreciation of reconstructed signal reliability vs. measurement device-(SPND - self powered neutron detectors -) outage-rate for nuclear safety purposes (IRSN - Institut de Radioprotection et de Sûreté Nucléaire -), and several other applications for industrial problems are underway and still pending.

Instead of “creating” information through any process, the SCM's original methodology propagates it from the point in the phase-space where it is experimentally available to the neighboring points - where it is either missing or false -. That is done adopting simple rules of degradation, which allow the information content decrease continuously with the distance from the “source” point (actually, the point where information is available), as it is the case for the force in the gravitational and electromagnetic fields.

The book is structured in three parts, which can be addressed quite independently from each others, depending on the reader's interest, taste and knowledge:

- The first part - which adopts a simple and intuitive approach - describes the ground-level construction rules of the SCM's methodology and provides the reader with the necessary background to understand it without any major theoretical effort. It does not require any specific advanced knowledge and practice of probability theory,

- The second one describes in some details the past practical applications of the SCM's method and provides with a critical insight on the results. It is of large interest for anybody who is interested by the problems of data completeness and reliability for practical use and purposes,

- Eventually, the third one - which is intended and reserved to a specialist's reading – provides the overall and comprehensive theoretical approach which supports the method.

Probabilistic Information Transfer, by Olga Zeydina & Bernard Beauzamy turns-out the achievement of a long questioning and the accomplished result of a long-way development, which started more than ten years ago and has, now, reached its full maturity.

The book is likely to become a reference for everybody who is interested in data completion and reconstruction problems for scientific and industrial purposes.

Moreover it is readable and quite accessible to not specialists - at least as far as the first two parts are concerned -. A very nice book, to read and to share.

Giovanni Bruna