



Probabilistic methods

Robust methods

Quite often, one has to take a decision without having all necessary information. This is true for organizational tasks (deliveries, planning), prospective (evolution of markets, of consumptions, and so on). In other circumstances, the existing data are very poor and often imprecise. In environmental questions, the physical laws are poorly known.

To replace the missing data by fictitious ones is certainly not a good solution: one computes for hours, and finally obtains a solution, the validity of which is by no means certain: the result will depend upon the data, and no confidence interval is known for them.

Probabilistic methods allow a first analysis, which leads to a hierarchy, in terms of orders of magnitude: one risk deserves to be taken into account, whereas some others may be neglected. Then, according to the needs, one can go further, and ask for a better knowledge of those which have been kept. In this respect, we realized in 2005 a risk analysis for the French "Commissariat à l'Energie Atomique", Saclay center. It dealt with a comparison and classification of risks linked with planes which may fly above the zone, and linked with transportation of dangerous materials, by trucks, near the site. The output of our study was a hierarchy of these risks.

When one uses probabilistic methods, the underlying phenomenon does not need to be connected with "randomness". In fact, in real life, nothing is really "random". But we decide to treat it this way, that is we consider that the risk is of random nature. One decides not to look for the precise causes of the phenomenon. The essential reason is that, in most cases, these precise causes depend upon physical or economical laws, which are poorly understood, and on which there are few data. Take the example of an electric bulb: you can try to understand the reasons why after some time it stops working, or you can simply decide that its life duration is a random variable.

During the years 2005-2006, in the framework of a contract with Veolia Environnement, West Region, we reconstructed the daily flows of 19 rivers in Vendée: they had been

measured for 37 years, but with 50 % of missing data. We used probabilistic methods (see the book by Bernard Beauzamy and Olga Zeydina: [RDM], reference below). This reconstruction was coarse, but precise enough to get an answer to the question: what is the importance of the lack of water during the summer, and how to solve it ? Do we need more dams ?

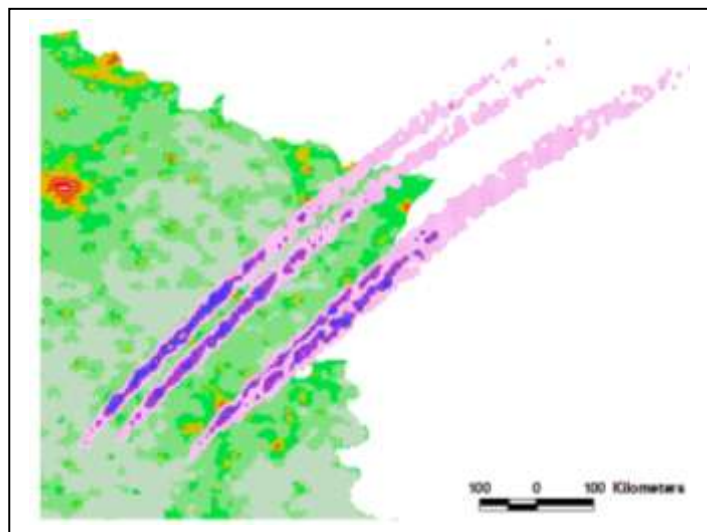
1. A simple and robust principle

The principle behind probabilistic methods is simple: we consider that everything that we do not know precisely depends on a probability law. This can be:

- Some data which are not precise. One would say, for instance, that natural gas consumption in France is linked with temperature, but not in a deterministic way. For a given temperature, many consumptions are possible.
- The law may be poorly understood. In 2004-2005, the French "Centre National d'Études Spatiales" asked us to study the risks associated with the debris, coming from the reentry of satellites. In the formula giving air resistance, we considered air density, at various altitudes, as a random variable, and the same for the exponent of the speed. Nobody is really sure that the air resistance is proportional to the square of the speed, for objects at a speed of 7 km/s, in the high atmosphere (100 km).

The result was not a precise point for the fall of a debris, but a probabilistic map: here is where the debris may fall, and with what probability. Such a map is given on this image. It corresponds to the fall of several debris, coming from one satellite, on the East of France.

The dark zones have higher probability to receive an object than the bright zones.



2. Improving the measurements

Probabilistic methods allow an improvement of the measurements, using proper "calibration tables". They are tables of conditional probabilities. They read the following way: assuming that a sensor gives this information, here is the probability of a given error. These laws are usually different along the whole scale of measure: a sensor is usually less precise at both ends. We used these methods in contracts with the French Ministry of Defense (improving the precision of a missile) and then with the Institut de Radioprotection et de Sûreté Nucléaire, 2003-2010: improving the precision of nuclear measurements.

3. Constructing predictive indicators

Probabilistic methods allow the construction of predictive indicators in a very robust manner. We already realized such indicators on three occasions: predictions of the prize of wheat, of cars sales, of Nickel stocks and uses. The idea is simply to find, in large databases, existing indicators which show a high degree of dependence (not linear correlation), after some time shift, with the quantity we want to predict.

4. Environmental concerns

Nowadays, many Companies need to control their emissions, for instance of CO₂, but also of many pollutants. But how to take into account the uncertainties upon the levels of emission ? For the so-called "CO₂ footprint", we showed that the classical approach contains severe methodological errors, and we showed, taking these errors into account, how to reduce the footprint.

5. What is the difference with statistics ?

Statistics allow many treatments upon data: adjustments, regressions, tests, and so on, but they always require an underlying assumption upon the law (for instance the law is assumed to be Gaussian, Poisson, and so on). However, in the situations where we are working, these laws are not known. To make an assumption upon them and to introduce artificial laws is not acceptable: it would be just as faulty as introducing artificial data.

Probabilistic methods do not have this drawback, since our work is primarily to build the probability law: see the example of gas consumption above. This law is drawn from existing data, no matter whether they are numerous or not. The law may be coarse, it is never artificial.

One might roughly say that statistics are a refined form of probabilities, when the laws are known, and that probabilities are a preliminary form of statistics, before the law is known. Of course, if historical data are abundant (sales of some good, for instance), there is no reason not to use the usual statistical tools.

6. Robust modeling

Probabilistic methods are an essential part of our research program "robust mathematical modeling", developed with many institutions, universities and companies, in France and abroad (see our web site <http://www.scmsa.eu/robust.htm> for a complete description of the program and a list of participating institutions).

A "robust" method takes into account, from the very beginning of the program, all uncertainties upon the laws and upon the data, but also upon the objectives. Indeed, experience has shown to us that an industrial project never has only one objective: there are short term and long term objectives, production problems, organization problems, human resources problems, and so on. The whole project is always complex; to reduce it to the

satisfaction of a single optimization criterion (often dealing with costs) is usually wrong, and leads to improper solutions.

So we consider that one should abandon the search for a precise optimum, and treat the whole sets of objectives as constraints. For instance: "spend 10% less than last year, reduce the backlog by 5%", and so on. This must be obtained quickly. This is what we call "*Quick Acceptable Solution*". When it has been found, one may start again, in order to improve it, with more precise constraints.

The idea is here that people in charge of the decisions will prefer a coarse and robust solution, obtained quickly, and allowing to know easily what are the favorable configurations, rather than a precise answer, requiring hours of computation, and based upon artificial or false data.

7. Industrial processes

Many people think that an industrial process is totally deterministic. But this is not true at all: the properties of the output usually have some variations, which are considered as unwanted by the industry. Many parameters may influence these variations: temperatures, pressures, chemical compositions, and so on. Our probabilistic methods allow to rank all these parameters in a "hierarchy": which ones are most influential upon the variability of the output? It means that, in order to reduce this variability, one will have to control first the parameters with high ranking.

For each parameter, we compute two conditional probability laws, for the output variable: one law when the parameter is small (lower than the median) and one law when it is large (larger than the median). We compute the difference between these two laws. The numbers obtained by this procedure allow the ranking. For more details, we refer to the book by Bernard Beauzamy [NMP], chapter III. A simple example was realized by SCM for the Water Agency "Agence de l'Eau Artois-Picardie", during the year 2008, and can be consulted: http://scmsa.eu/archives/SCM_AEAP_2008_12_01.pdf

8. Looking for dangerous zones

This comes as a complement of the previous paragraph. We have a process, depending on a large number of parameters, and we want to find the configurations of the parameters which will lead to a "risk zone", for the output variable of the process. This risk zone might be for instance a quality which is too low, or a temperature which is too high. Our methods allow the description of such risk zones, only from the existing data, and with no extra assumption (for instance, we do not have to assume that the model is linear, or Gaussian, or anything else). This method is described in the chapter XIII of the book [NMP].

9. Archimedes' Method

It was considered by Archimedes as its masterpiece, and was lost during more than 2000 years. We just started exploiting it. The general idea is to compare the existing information to a known information, generated for this purpose. Such a comparison is much more robust than the formulas which are generally used for the computations. Archimedes' Weighing Method leads to astonishing projects ; see the book [AMW], reference below.

Books:

[MPPR] Bernard Beauzamy: Méthodes probabilistes pour l'étude des phénomènes réels, ISBN: 2-9521458-0-6, Editions de la SCM, March 2004 (in French). Second edition, June 2016.

[RDM] Bernard Beauzamy et Olga Zeydina: Méthodes probabilistes pour la reconstruction de données manquantes, ISBN: 2-9521458-2-2, Editions de la SCM, April 2007 (in French).

[NMP] Bernard Beauzamy: Nouvelles méthodes probabilistes pour l'évaluation des risques. ISBN: 978-2-9521458-4-8, ISSN: 1767-1175, Editions de la SCM, April 2010 (in French).

[AMW] Bernard Beauzamy: Archimedes' Modern Works. ISBN : 978-2-9521458-7-9, ISSN : 1767-1175. Editions de la SCM, August 2012 (in English).

[PIT] Olga Zeydina and Bernard Beauzamy : Probabilistic Information Transfer. ISBN: 978-2-9521458-6-2, ISSN: 1767-1175. Editions de la SCM, mai 2013.

Recent references:

- IRSN, 2006-2008: Probabilistic methods for nuclear safety: definition of the Experimental Probabilistic Hypersurface (a method created by SCM).
- Direction Générale de l'Energie et des Matières Premières (French Ministry of Finances): Probabilistic study on the risks associated with the imports of natural gas.
- French National Agency for Nuclear Waste (ANDRA), 2007-2008: Probabilistic analysis of models of transfers for radionuclides.
- European Environment Agency, since 2006: General probabilistic methods for the environment, with specific application to the quality of the water in rivers.
- CEA, Direction de l'Energie Nucléaire: Probabilistic methods in seismology, 2007, in epidemiology, 2007-2008.
- Delegation for Nuclear Safety for Defense Installations (DSND), 2007-2008: Probabilistic risk assessment for nuclear weapons systems.
- Sodebo, 2008-2009: Construction of a prospective indicator about wheat prices.
- Caisse Centrale de Réassurance, 2009: Probabilistic Methods for river flows.
- SNECMA Propulsion Solide, 2009-2010: Probabilistic Methods for reliability.
- Rhodia, 2009: Construction of a prospective indicator about worldwide car sales.
- International Stainless Steel Forum, 2010: Analysis of CO2 emissions.

- Nuclear Energy Agency (OCDE), 2010: Probabilistic methods for the detection of erroneous data in large databases.
- Groupe Total, 2010: Probabilistic methods for the evaluation of the amount of pollutant.
- Caisse Centrale de Réassurance, 2010-2011: Probabilistic methods for the evaluation of extreme phenomena.
- Direction Générale Energie Climat, 2010-2011, with CITEPA: Estimates of the uncertainties in a national inventory for pollutants.
- PSA Peugeot Citroën, 2011: Probabilistic studies for the extension of warranties for the cars.
- Réseau Ferré de France, 2011: Probabilistic studies related to the delays of the trains in the Paris region.
- IRSN, 2011: Probabilistic methods for nuclear safety.
- International Stainless Steel Forum, 2011: Probabilistic tools for the forecast of Nickel prices and sales.
- Commission Européenne (with the Poyry Group), 2011-2012: Probabilistic methods for water quality.
- IFSTTAR, 2011-2014: Probabilistic methods for precise positioning, using GPS in an urban environment.
- Suez Environnement, 2011-2012: Probabilistic methods for water quality.
- ArcelorMittal, 2011-2012: Probabilistic methods for the quality of an industrial process.
- Nuclear Energy Agency (OCDE), 2011-2012, 2014: detection of erroneous data in large databases, using probabilistic techniques.
- Air Liquide, 2011 : Hierarchy of parameters and construction of a similarity index between pipelines.
- Areva, 2013 : Hierarchy of parameters in an industrial process.
- DCNS, Indret, 2013 : Hierarchy of parameters in an industrial process.
- Coop de France déshydratation, 2013 : Hierarchy of parameters and their influence upon a deshydration process.
- IRSN, 2014-2015 : Critical analysis of the TELERAY network (observation of radioactivity in the environment).
- Direction Générale Energie-Climat (French Ministry of Environment), Bureau Qualité de l'Air, 2015 : Probabilistic links between traffic and pollutants on the "boulevard périphérique" around Paris.
- ERDF (French Electricity distribution), 2015 : Critical analysis of the organization of the work in order to recollect the electricity consumptions.
- Solétanche-Bachy, 2015 : Hierarchy of parameters which may have an influence upon the deformation of a construction.
- Telcap, 2015 : previsions for traffic on phone lines.
- Carrefour, 2016 : Hierarchy of parameters which may have an influence upon the sales of some goods.
- COSEA, 2016 : Hierarchy of parameters which may have an influence upon water quality
- G7 (Taxis), 2016 : Statistical analysis.