Probabilistic methods

Robust methods

Quite often, one has to take a decision without having all the 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. This is typically the case about the environment, where the laws are moreover poorly known.

To replace the missing data by fictitious data is certainly not a good solution : one computes for hours, and finally one 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 center, and linked with transportation of dangerous materials, by trucks, near the site. The output of our study was a hierarchy of these risks.

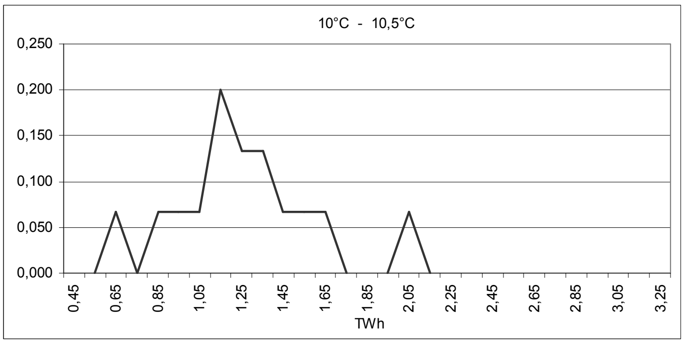
When one uses probabilistic methods, the underlying phenomenon does not need to be connected with "randomness". In fact, in real life, nothing is really "random". But we decide to treat it this way, that is to decide that the risk is of random nature. One decides not to look for the precise causes of the phenomenon. The essential reason is that, most of the cases, these precise causes depend upon physical 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 frame 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 : Méthodes probabilistes pour la reconstruction de données manquantes, Editions de la SCM, 2007). 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 ?

A simple and robust principle

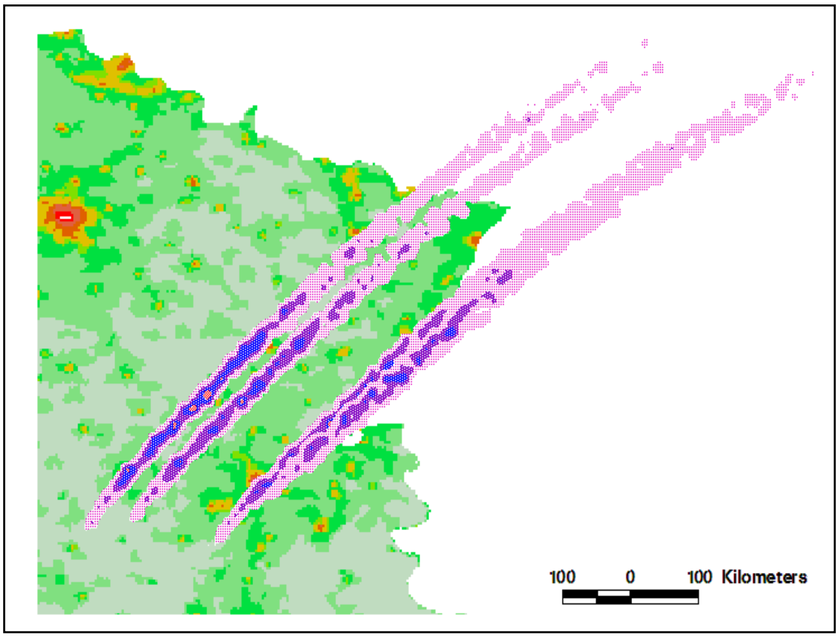
The principle behind probabilistic methods is simple : we consider that everything that we do not know precisely depends on a probability law. This can be :

– 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.



The graph above comes from a study we did in 2006 for the French "Direction Générale de l'Energie et des Matières Premières" (Ministry of Finances) : For a given temperature (in the interval 10°C - 10.5°C), the x-axis is a consumption (in TWh) and the y-axis is the probability of this consumption. Several consumptions are possible, not all with the same probability.

– The law may be poorly understood. In 2004-2005, the French "Centre National d'Etudes Spatiales" asked us to study the risks associated with the debris, coming from the reentry of satellites. Data are not very numerous, and the laws are poorly understood. In particular, in the formula giving air resistance, we considered that 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 the image, left. It corresponds to the fall of several debris, coming from one satellite.

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

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, 2003-2007 : improving the precision of nuclear measurements.

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). But, 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.

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.com/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 bring 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.

Books :

Bernard Beauzamy : Méthodes probabilistes pour l'étude des phénomènes réels, ISBN : 2-9521458-0-6, Editions de la SCM, mars 2004.

Bernard Beauzamy et Olga Zeydina : Méthodes probabilistes pour la reconstruction de données manquantes,ISBN : 2-9521458-2-2, Editions de la SCM, avril 2007.

Recent references :

- INERIS, 1999-2002 : Mathematical modeling of toxic effects on living species.
- Institut de Protection et de Sûreté Nucléaire (IPSN), 2000 : Sensitivity analyses on numerical codes.
- Framatome-ANP, 2003-2004 : Statistical methods on thermo-hydraulic analyses in the accidental studies for nuclear reactors.
- Centre National d'Etudes Spatiales, 2003 : Analysis of the method used in order to determine the risks connected with the fall of a balloon.
- Institut de Radioprotection et de Sûreté Nucléaire, 2003-2007 : Improving the measurements of Uranium and Plutonium, using probabilistic methods.
- Veolia Environnement, 2003-2005 : Prospective analysis related to the used water systems for the city of Brest, for the years 2010-2015.
- Centre National d'Etudes Spatiales, 2004-2005 : Probabilistic maps connected with the reentry of space debris.
- French Electricity : Management of the production, for a 8 days prevision.
- French Railways (SNCF), 2005-2006 : Probabilistic studies for the "Compression longitudinal efforts", with application to the length of freight trains.
- Commissariat à l'Energie Atomique, Saclay site, 2005-2006 : Probabilistic analysis of the industrial risks connected to planes flying near the site and to trucks carrying dangerous goods.
- Veolia Environnement, 2005-2006 : Probabilistic methods in order to characterize the lack of water in Vendée.
- Veolia Transport (Connex), 2005-2006 : Probabilistic methods for the definition of public transportation nets.
- Espaces Ferroviaires, 2006 : Analysis of the risks connected with real estate operations.
- IRSN, 2006-2007 : 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 concerning the risks associated with French imports of natural gas.
- Agence Nationale pour la Gestion des Déchets Radioactifs (ANDRA), 2007 : Probabilistic analysis of models of transfers for radionuclides.
- European Environment Agency, since 2006 : General probabilistic methods for environment.
- CEA, Direction de l'Energie Nucléaire, Département Modélisation des Systèmes et Structures : Probabilistic methods in seismology, 2007.