Somme comments on the contents of the RMM site

The basic motivation seems undeniable, at least to someone who has been once confronted with a complex, "real life" problem and has had the aim of giving it a solution that can be reasonably implemented. The probabilistic methods the are described are certainly interesting and remind one of the empirical Bayes "philosophy" (due to Herberrt Robbins, the 50's) wich consists in providing likelihood intervals that locate the quantities of interest and uses the observations that are available to "calibrate" the likelihood functions one uses. Empirical Bayes techniques seem nowdays very popular among analyists of microarray structures exactly for the reasons given in the RMM arguments : low data to parameters ratios. It nevertheless remains a fact that any technique is more or less appropriate and more or less efficient depending on the problem that is attempted and the context in which this problem lives. Thus when optimizing entropy one is often provided with a solution in the form a Gaussian law (perhaps because entropy involves the logarithm) which to many may appear as highly unlikely. As one is unlikely to master all possible ways to answer a question or solve a problem, a realistic and reasonable position (provided one has worked diligently and honestly) is to say : "Here are my conclusions and I have such and such confidence(diffidence) that these conclusions obtain, for such and such reasons."

To illustrate the latter point I shall provide an example. I was once asked to evaluate the ability of a piece of apparatus to detect glaucomas (including those without pressure). The apparatus was based on the use of a physiological curve whose shape was supposed to yield the type of glaucoma involved (none was also a type !). The multicentric study that had been carried out to evaluate the apparatus's efficiency had been so poorly done that none of the indicators the curve was supposed to provide had statistical "clearance." However, while studying those curves, I noticed that the area delimited by each curve seeemed to play a part. I did a quick and dirty evaluation of these areas for all the cases and, lo and behold, the area seemed to do the trick. I reported that there seemed to be something worth pursuing in that apparatus, though not what was originally believed to be there, but that, to have some certainty that the phenomenon could be ascertained and controlled, one had to repeat the multicentric study and control it much more tightly. Do I need to say that the R&D budget had already been depleted and that there was no more money available ? So the apparatus never saw the light of day !

One "caveat :" customers seldom like to be told that there is unavoidable uncertainty in the answers!

One surprise in the site's contents : the absence of some of the robust wares that are already on the market, if only to say that they are not fit for service. I know of and have some experience with three areas in which robust concerns predominate. Here are a few comments on each.

Robust statistics

Three names come to mind when one summons robust statistics : Tukey, Huber and Hampel (THH). The original concern of robust statistics was the stability and representativeness of the mean and the standard deviation : when nine customers spend one thousand and one spends one million, the mean expense neither represents the low spenders nor the high spender. The old times answer was trimming the outlier. But THH provided general frameworks within which to think statistical robustness. Huber thus suggested one uses minimax in a neighbourhood of the hypothesized model while Hampel required continuity with respect to convergence in law. For estimation problems Huber introduced functions that downweight large outliers (a form of trimming : the "best" data sets seem to contain percentages of errors that cannot be overlooked). Tukey is the father of the jackknife that fits the same model with a family of data sets, those obtained from the available observations by leaving out a fixed percentage of these. One thus gets an idea of the model's stability.

While robust statistics was for a while "all the rage," the wave has somewhat receded. There are several possible reasons for that. The first is that some think that it is a mistake to eliminate or downweight outliers as they are intrinsic to the measurement process. Thus the US Bureau of Standards is adamant against trimming outliers in its weight calibration studies where, for each hundred observations, there are three to five that are clearly outliers (see the text of Purves, Pisani at al.). Beyond such considerations is the practical fact that the implementation of robust methods requires hard decisions about the amount of trimming for example, or the form of the downweighting function that are pretty much as arbitrary as any other one makes in deciding which model to retain. The results of such decisions are computationally pretty much opaque procedures whose end effects are difficult to evaluate. And that leads to the third category of reasons : recourse to robust statistical methods tends to increase the heterogeneity of tools that enter an analysis. That makes the reliability assessment that was hinted at in the first paragraph above more difficult than when using standard methods which enjoy a greater degree of homogeneity as they have a "unique" origin, the conjunction of least squares with the Gaussian law.

To give an example, I once studied some fiscal data relating to the twenty-six states of Switzerland. With respect to such questions Switzerland is extraordinarily heterogeneous and I was pretty happy to succeed, by using robust regression techniques, in being able to say something not entirely trivial about the meaning of the data. But I felt that no statistical inference was warranted as to get to the end of the analysis I had to perform too many "statistical contortions." A paper on the topic was submitted to a Swiss statistics and economics journal and was turned down as the editor could not accept that I would not perfoms significance tests.

Robustness in analytical chemistry

It is the area in which the ideas about robustness seem to me the closest to those populating the RMM site though the techniques used are more of the usual statistical kind than those say of experimental probabilistic hypersurfaces. The problem chemists have is as follows. They develop methods to separate the compounds of a product and these methods often require that precise values of parameters such as temperature be maintained. They know however that when these methods are "exported" on a large scale, it will be difficult to obtain and maintain these critical values. The aim of robustness analyses is then that of predicting how far off the results will be when the required values do not obtain and from there to decide what measures to take to be in conformity with protocols.

Robustness in sonar

It is an area where one does what one can rather than what one wants! When one consults a "classic" text (for exemple : M. Bouvet, Traitement des signaux pour les systèmes sonar, Masson) one basically finds linear, stationary and Gaussian models that are rather atypical of what takes place in the ocean. But even in that context there arise robustness questions of a new kind. Indeed sound signals are realizations of stochastic processes and the associated statistical problems must then be framed in a infinite-dimensional world. Thus existence of a likelihood ratio (which means in particular that detection cannot take place without error) is not guaranteed and the conditions that allow one to use that likelihood are of a very precise nature. In the Gaussain case for example, the covariance operators of respectively the noise and the signal corrupted by that noise must be related by a Hilbert-Schmidt operator whose eigenvalues are strictly greater than one. How does one come to terms with such requirements when the number of observed points, hoever large, is nevertheless finite (de Brucq : sampling signals every $\Delta t = 1,2510^{-4}$ seconds, to study frequencies in the range of 0 à 3000 Hertz, yields, per hour, about 10^7 data points) ?

One solution I have become familiar with has the following characteristics. Because of the analytical difficulties of turning one's back to Gaussian laws, one postulates a noise model of the form AG, A positive and independent of the Gaussian G : that yields a rather general class of noise models (since one thus postulates mixtures of Gaussians), for instance models with impulsive noise. One makes no distributional assumptions on the signal as it is a transient and that there is no hope of ascertaining its distributional properties. One requires only that it has enough smoothness for the problem to be non-singular (that is have a likelihood). Such a procedure yields a generic, fairly general likelihood function that then gets calibrated (using painstakingly lengthy empirical calibration exercises). I am told that in certain situations, for small false alarm probabilities, one can get substantial improvements over existing methods.

Here are two final remarks. Sometimes the need for robust methods can be circumvented.

One example is as follows. Mandatory health insurance is, in Switzerland, managed by private insurance companies. These strive to gather "good risks" (the young) and avoid "bad risks" (the old). To restrain such practices there is an equalization mechanism which forces companies with "too many good risks" to release some of their income to the companies that have "too many bad risks." The "too many" is based on age and sex. It is very easy to convince oneself that sex and age are fram from sufficient to delineate the actual costs of health insurance. In an effort to better manage there has been, in Switzerland, a national research project with the aim to produce at least one more criterion that would improve the equalizing mechanism. The project has been carried mostly by health economists and econometricians. Unsurprisingly it has been a regression exercise (logit and probit) whose reliability is hard to fathom given the complexity of the subject matter and the assumptions that have been necessary to complete the calculations. A wiser approach consists in using as statistical unit the cumulated reimbursements for health care that the insured receive : health conditions are obvious and no assumptions are are required, only a good data base and efficient data handling capabilities (there would be about seven million health histories to deal with).

Another case is financial modelling. Nobody in finance believes, except perhaps over-enthusiastic beginners, that the sophisticated mathematical models that are in use have much bearing on reality (The Gaussian is bad, why not use the t-distribution ? It has long tails.). But little does it matter as long as the financial community at large uses the same models : what is needed it seems is a unique answer rather than the right answer. If everybody prices an option using Black-Scholes formula little does it matter what the actual price is : the price IS the result of the calculation. Indeed when I once tried to convince a fund manager to use what appeared to me more sensible methods that those that were, I was told in no uncertain terms that that would be very bad for business as customers would be very suspicious of investing methods nobody else would practice !