Preface

The book of nature, according to Galilei, is written in the language of mathematics. The nature of mathematics is being exact, and its exactness is underlined by the formalism used by mathematicians to write it. This formalism, characterized by theorems and proofs, and syncopated with occasional lemmas, remarks and corollaries, is so deeply ingrained that mathematicians feel uncomfortable when the pattern is broken, to the point of giving the impression that the attitude of mathematicians towards the way mathematics should be written is almost moralistic. There is a definition often quoted, "A mathematician is a person who proves theorems", and a similar, more alchemistic one, credited to Paul Erdős, but more likely going back to Alfréd Rényi, stating that "A mathematician is a machine that transforms coffee into theorems[1]". Therefore it seems to be the form, not the content, that characterizes mathematics, similarly to what happens in any formal moralistic code wherein form takes precedence over content.

This book is deliberately written in a very different manner, without a single theorem or proof. Since morality has its subjective component, to paraphrase Manuel Vasquez Montalban, we could call it *Ten Immoral Mathematical Recipes*[2]. Does the lack of theorems and proofs mean that the book is more inaccurate than traditional books of mathematics? Or is it possibly just a sign of lack of coffee? This is our first open question.

Exactness is an interesting concept. Italo Calvino, in his *Lezioni Americane*[3], listed exactness as one of the values that he would have wanted to take along to the 21st century. Exactness, for Calvino, meant precise linguistic ex-

[1] That said, academic mathematics departments should invest on high quality coffee beans and decent coffee makers, in hope of better theorems. As Paul Turán, a third Hungarian mathematician, remarked, "weak coffee is fit only to produce lemmas".

[2] M. V. Montalban: *Ricette immorali* (orig. *Las recetas inmorales*, 1981), Feltrinelli, 1992.

[3] I. Calvino: *Lezioni Americane*, Oscar Mondadori, 1988.

pression, but in a particular sense. To explain what he meant by exactness, he used a surprising example of exact expression: the poetry of Giacomo Leopardi, with all its ambiguities and suggestive images. According to Calvino, when obsessed with a formal language that is void of ambiguities, one loses the capability of expressing emotions exactly, while by liberating the language and making it vague, one creates space for the most exact of all expressions, poetry. Thus, the exactness of expression is beyond the language. We feel the same way about mathematics.

Mathematics is a wonderful tool to express liberally such concepts as qualitative subjective beliefs, but by trying to formalize too strictly how to express them, we may end up creating beautiful mathematics that has a life of its own, in its own academic environment, but which is completely estranged to what we initially set forth. The goal of this book is to show how to solve problems instead of proving theorems. This mischievous and somewhat provocative statement should be understood in the spirit of Peter Lax' comment in an interview given on the occasion of his receiving the 2005 Abel Prize[4]: "When a mathematician says he has solved the problem he means he knows the solution exists, that it's unique, but very often not much more." Going through mathematical proofs is a serious piece of work: we hope that reading this book feels less like work and more like a thought-provoking experience.

The statistical interpretation, and in particular the Bayesian point of view, plays a central role in this book. Why is it so important to emphasize the philosophical difference between statistical and non-statistical approaches to modelling and problem solving? There are two compelling reasons.

The first one is very practical: admitting the lack of information by modelling the unknown parameters as random variables and encoding the nature of uncertainty into probability densities gives a great freedom to develop the models without having to worry too much about whether solutions exist or are unique. The solution in Bayesian statistics, in fact, is not a single value of the unknowns, but a probability distribution of possible values, that always exists. Moreover, there are often pieces of qualitative information available that simply do not yield to classical methods, but which have a natural interpretation in the Bayesian framework.

It is often claimed, in particular by mathematician in inverse problems working with classical regularization methods, that the Bayesian approach is yet another way of introducing regularization into problems where the data are insufficient or of low quality, and that every prior can be replaced by an appropriately chosen penalty. Such statement may seem correct in particular cases when limited computational resources and lack of time force one to use the Bayesian techniques for finding a single value, typically the maximum a posteriori estimate, but in general the claim is wrong. The Bayesian framework, as we shall reiterate over and again in this book, can be used to produce

[4] M. Raussen and C. Skau: *Interview with Peter D. Lax.* Notices of the AMS **53** (2006) 223–229.

particular estimators that coincide with classical regularized solutions, but the framework itself does not reduce to these solutions, and claiming so would be an abuse of syllogism[5].

The second, more compelling reason for advocating the Bayesian approach, has to do with the interpretation of mathematical models. It is well understood, and generally accepted, that a computational model is always a simplification. As George E. P. Box noted, "all models are wrong, some are useful". As computational capabilities have grown, an urge to enrich existing models with new details has emerged. This is particularly true in areas like computational systems biology, where the new paradigm is to study the joint effect of a huge number of details[6] rather than using the reductionist approach and seeking simplified lumped models whose behavior would be well understood. As a consequence, the computational models contain so many model parameters that hoping to determine them based on few observations is simply unreasonable. In the old paradigm, one could say that there are some values of the model parameters that correspond in an "optimal way" to what can be observed. The identifiability of a model by idealized data is a classic topic of research in applied mathematics. From the old paradigm, we have also inherited the faith in the power of single outputs. Given a simplistic electrophysiological model of the heart, a physician would want to see the simulated electrocardiogram. If the model was simple, for example two rotating dipoles, that output would be about all the model could produce, and no big surprises were to be expected. Likewise, given a model for millions of neurons, the physician would want to see a simulated cerebral response to a stimulus. But here is the big difference: the complex model, unlike the simple dipole model, can produce a continuum of outputs corresponding to fictitious data, never measured by anybody in the past or the future. The validity of the model is assessed according to whether the simulated output corresponds to what the physician *expects*. While when modelling the heart by a few dipoles, a single simulated output could still make sense, in the second case the situation is much more complicated. Since the model is overparametrized, the system cannot be identified by available or even hypothetical data and it is possible to obtain completely different outputs simply by adjusting the parameters. This observation can lead researchers to state, in frustration, "well, you can make your model do whatever you want, so what's the point"[7]. This sense of hopelessness is exactly what the Bayesian approach seeks to remove. Suppose that the values of the parameters in the

[5] A classic example of analogous abuse of logic can be found in elementary books of logic: while it is true that Aristotle is a Greek, it is not true that a Greek is Aristotle.

[6] This principle is often referred to as *emergence*, as new unforeseen and qualitatively different features emerge as a sum of its parts (cf. physics $\longrightarrow$ chemistry $\longrightarrow$ life $\longrightarrow$ intelligence). Needless to say, this holistic principle is old, and can be traced back to ancient philosophers.

[7] We have actually heard this type of statement repeatedly from people who refuse to consider the Bayesian approach to problem solving.

complex model have been set so that the simulated output is completely in conflict with what the experts expect. The reaction to such output would be to think that the current settings of the parameters "must" be wrong, and there would usually be unanimous consensus about the incorrectness of the model prediction. This situation clearly demonstrates that some combinations of the parameter values have to be excluded, and the exclusion principle is based on the observed data (likelihood), or in lack thereof, on the subjective belief of an expert (prior). Thanks to this exclusion, the model can no longer do whatever we want, yet we have not reduced its complexity and thereby its capacity to capture complex, unforeseen, but possible, phenomena. By following the principle of exclusion and subjective learned opinions, we effectively narrow down the probability distributions of the model parameters so that the model produced plausible results. This process is cumulative: when new information arrives, old information is not rejected, as is often the case in the infamous "model fitting by parameter tweaking", but included as prior information. This mode of building models is not only Bayesian, but also Popperian[8] in the wide sense: data is used to *falsify* hypotheses thus leading to the removal of impossible events or to assigning them as unlikely, rather than to verify hypotheses, which is in itself a dubious project. As the classic philosophic argument goes, producing one white swan, or, for that matter, three, does not prove the theory that all swans are white. Unfortunately, deterministic models are often used in this way: one, or three, successful reconstructions are shown as proof of a concept.

The statistical nature of parameters in complex models serves also another purpose. When writing a complex model for the brain, for instance, we expect that the model is, at least to some extent, generic and representative, and thus capable of explaining not one but a whole population of brains. To our grace, or disgrace, not all brains are equal. Therefore, even without a reference to the subjective nature of information, a statistical model simply admits the diversity of those obscure objects of our modelling desires.

This books, which is based on notes for courses that we taught at Case Western Reserve University, Helsinki University of Technology, and at the University of Udine, is a tutorial rather than an in-depth treatise in Bayesian statistics, scientific computing and inverse problems. When compiling the bibliography, we faced the difficult decision of what to include and what to leave out. Being at the crossroad of three mature branches of research, statistics, numerical analysis and inverse problems, we were faced with three vast horizons, as there were three times as many people whose contributions should have been acknowledged. Since compiling a comprehensive bibliography seemed a herculean task, in the end Occam's razor won and we opted to list only the books that were suggested to our brave students, whom we thank for feedback and comments. We also want to thank Dario Fasino for his great hos-

---

[8] See A. Tarantola: *Inverse problems, Popper and Bayes*, Nature Physics **2** 492–494,(2006).

pitality during our visit to Udine, Rebecca Calvetti, Rachael Hageman and Rossana Occhipinti for help with proofreading. The financial support of the Finnish Cultural Foundation for Erkki Somersalo during the completion of the manuscript is gratefully acknowledged.

Cleveland – Helsinki, *Daniela Calvetti*
May 2007 *Erkki Somersalo*