Popper, Bayes and the inverse problem

ALBERT TARANTOLA

is at the Institut de Physique du Globe de Paris, 4 place Jussieu, 75005 Paris, France.

e-mail: tarantola@ipgp.org

Using observations to infer the values of some parameters corresponds to solving an 'inverse problem'. Practitioners usually seek the 'best solution' implied by the data, but observations should only be used to falsify possible solutions, not to deduce any particular solution.

As we learned from Popper[1], a theory must be able to predict the result of observations. And although any number of successful predictions cannot guarantee that a theory is correct, one single wrong prediction falsifies the theory, thus requiring its modification or, sometimes, its replacement by a better one. Using a physical theory for predicting the results of observations corresponds to solving the 'forward modelling problem'. The reciprocal situation, using the result of measurements to infer the values of the parameters representing a system, corresponds to the 'inverse modelling problem'.

Whereas the forward problem has (in non-quantum physics) a unique solution, because of the causality principle, the inverse problem may have many solutions — when different models of the system predict similar observations, or no solution at all (inconsistent data). This is why a special mathematical theory exists that formalizes inverse problems.

FORMALIZATION OF THE INVERSE PROBLEM

The inference of the values of some parameters based on observations is, of course, as old as quantitative science, but it was only during 1760–1810 that the first formalizations arose[2]. The two basic problems of that time were the use of geodetic data to estimate the shape of the Earth, and the use of astronomical data to infer the orbits of planets and comets.

The problem was then stated as an optimization problem: which values of the model parameters best fit the observations? While some scientists (Boscovich and Laplace) were minimizing the sum of the absolute values of the misfits — the least-absolute-values method — other scientists (Legendre and Gauss) were minimizing the sum of the squared values of the misfits — the least-squares method. It soon became clear that each of the two methods followed from a different hypothesis about the statistical distribution of errors in observations, the laplacian distribution $f(x) \sim \exp(-|x|)$ requiring the use of the least-absolute-values method, and the gaussian distribution $f(x) \sim \exp(-x^2)$ requiring the use of the least-squares method.

Today, the overwhelming popularity of the least-squares method, with respect to the least-absolute-values method, is not due to any superior performance in interpreting typical data sets, but to the fact that solving a least-squares problem only involves the use of simple linear algebra; solving a least-absolute-values problem requires more complex computations (such as linear programming). It is widely known that the least-absolute-values criterion is much less sensitive than least squares to the presence of large uncontrolled errors in the data (a property called robustness).

The International Organization for Standardization (ISO) is clear in recommending[3] the expression of the uncertainty in the result of a

measurement not as something like $x = x_0 \pm s$, but, more generally, as a probability distribution $f(x)$ over the quantity x being measured. Whereas, as mentioned above, the laplacian and gaussian forms of $f(x)$ lead to the least-absolute-values criterion and the least-squares criterion, respectively, any other possible form of $f(x)$ implies its own minimization criterion.

THE NEED FOR DEEPER UNDERSTANDING

At the beginning of the 1970s, geophysicists made advances towards the understanding of the inverse problem, driven by an acute need in that field: geophysicists spend most of their time trying to infer the properties of planetary interiors using only data obtained at the surface. Although the modern theory was initiated by geophysicists, it is clear that the inverse problem can, and should, be applied in all the fields of physics. For instance, a typical inverse problem in astronomy is that of inferring the three-dimensional luminosity density of the Milky Way (Fig. 1) using data on the two-dimensional brightness observed on the sky sphere. Also, the interpretation of Rutherford's gold foil experiment, which explained the deflection of alpha particles by a model of electric charge distribution in the atom, could have been stated as an inverse problem (at least for the quantitative part of the problem).

When solving inverse problems, scientists may be faced with two very different difficulties. The first is to find at least one model of the system that is consistent with the observations. This is sometimes a non-trivial affair. In my area of geophysics, for instance, seismometers located on the slopes of a volcano (Fig. 2) record extremely complex seismograms, and all available models of the volcano predict seismograms that are hopelessly different from the observed ones. This problem is akin to that of finding a needle in a haystack. Although a relatively easy task in three dimensions, it is very hard if the haystack — the space of all possible models — has hundreds of dimensions. We will come back to this problem later.

The second difficulty is, in problems where finding at least one solution is doable, how can we quantify the non-uniqueness of the result? For instance, in the Milky Way problem mentioned above, how many different three-dimensional models of luminosity density may explain the observed brightness distribution? More dramatically, it is well known that there are infinitely many possible models of mass density inside a planet that can exactly fit any number of observations of the gravity field outside or at the surface of the planet.

SHOULD *A PRIORI* INFORMATION BE USED?

Two different philosophies are in use today. The first carefully avoids using any *a priori* information on the model parameters that could 'bias' the inferences to be drawn from the data. This approach is favoured, in particular, by statisticians, as it can easily be formalized and presented with standard mathematical rigour. The theory developed by Backus and Gilbert[4,5] is a good example of this approach.

Figure 1 **Sideways. An edge-on view of our galaxy, the Milky Way, taken by the Cosmic Background Explorer (COBE) satellite. How many models are consistent with the measured luminosity?**

The second philosophy is clearly bayesian, asking the basic question: how does the newly acquired data modify our previous information? In other words, when starting with some *a priori* state of information on the model parameter, which is the *a posteriori* state of information at which we arrive after 'assimilating' new data[6]? Although I think this second point of view is better adapted to deal with real-life uncertainties, the inevitable subjective content of the reasoning sparks fire from statisticians. In fact, I am going to present below an even more extreme point of view, by invoking both Bayes and Popper.

Let me first argue that the very idea of using observations to infer one model of the system (the 'best model' or the 'mean model' or whatever) is wrong. Observations cannot produce models, they can only falsify models. Of course, any displayed model should be accompanied by some estimation of uncertainties. But normally those uncertainties are either trivial, and can be described by the simple gaussian model, or impossible to manage — for instance, a severely multimodal probability distribution in a large-dimensional model parameter space.

FINDING A NEEDLE IN A HAYSTACK IS HARD IF THE HAYSTACK HAS HUNDREDS OF DIMENSIONS

There is also a psychological problem with the computation — and subsequent display — of this 'best solution': our cognitive system has the inclination, after some period, to attribute to this model of the system a veracity that the associated uncertainties (that are, usually, very difficult to grasp) do not warrant. This possibly results from a darwinian adaptation of animals: do not spend too much time understanding uncertainties; one model of the world must be adopted (a predator is approaching) and adequate action must be taken (run!). The solution to a problem based on an interpretation of observations should not consist of a display of one solution accompanied by a technical description of the associated uncertainties, but, rather, by a proper display of all (or as many as possible) solutions that are consistent with the observations. Let us see how this can be done — at least in principle.

WWW.ARENAL.NET

Figure 2 **Giant 'haystack'. Seismometers placed on volcanoes record complex seismograms, most of which do not agree with available models. The search for the needle in the haystack continues.**

PROGRESS BY FALSIFICATION

In the traditional bayesian approach, there is an explicit introduction of some *a priori* information on the system, in addition to the uncertainties attached to any measurement; when using gravity and seismology data to infer the mass density at the centre of the Earth, we use the *a priori* information that we are dealing with an iron alloy at high pressure, for which we have a reasonable *a priori* range for the mass density. For some simple problems, this approach just leads to an optimization algorithm that is in all points similar to those used by Laplace and Gauss, which provide both a solution (the 'best model') and an estimation of the errors (differences between our model and the 'true model').

But for many everyday problems, optimization techniques don't work because the posterior probability distribution in the model parameter space is complex; there could be many secondary optima, for example, or whole subspaces with complex geometry that may be totally equivalent. Consider, for instance, the typical inverse problem in oil exploration; the search for a model of a sedimentary basin uses some basic *a priori* tectonic information along with seismograms obtained from a seismic survey generated from artificial sources of seismic waves. The complexity of the medium (with faulted irregular layers, folds and so on) produces very complex seismograms. The challenge here is twofold: even if we suspect that there may be very many models that are consistent with the observations, finding a single one may be a daunting task because of strong nonlinearities in the functions to be optimized and the large dimensionality of the spaces involved.

The idea that solving the inverse problem corresponds to obtaining one 'best' model requires revision. All the literature based on this paradigm suggests that solving an inverse problem consists somehow in 'extracting' a model from the data, using techniques reminiscent of those in signal processing. This is not necessarily the right point of view. What we should do is to base the inverse problem on a modified version of the popperian paradigm: data are not to be used to create a model, but, instead, to falsify models.

I suggest that the setting, in principle, for an inverse problem should be as follows: use all available *a priori* information to sequentially create models of the system, potentially an infinite number of them. For each model, solve the forward modelling problem, compare the predictions to the actual observations and use some criterion to decide if the fit is acceptable or unacceptable, given the uncertainties in the observations and, perhaps, in the physical theory being used. The unacceptable models have been falsified, and must be dropped. The collection of all the models that have not been falsified represent the solution of the inverse problem.

This concept of passing from a 'prior collection of models' to a 'posterior collection of models' will certainly be acceptable by the lovers of Bayes' paradigm, as the collections of models can be seen as samples of a prior probability distribution and samples of a posterior distribution. It should also please the believers in Popper's point of view. Although still far from unanimity, there are quests in this world too difficult to attain.

REFERENCES

1. Popper, K. *The Logic of Scientific Discovery* (Basic Books, 1959).
2. Stigler, M. S. *The History of Statistics: The Measurement of Uncertainty Before 1900* (Belknap, Harvard, 1986).
3. *Guide to the Expression of Uncertainty in Measurement* (International Organization for Standardization, Switzerland, 1995).
4. Backus, G. *Proc. Natl Acad. Sci.* **65,** 1–105 (1970); *ibid.* **65,** 281–287 (1970); *ibid.* **67,** 282–289 (1970).
5. Backus, G. & Gilbert, F. *Philos. Trans. R. Soc. London* **266,** 123–192 (1970).
6. Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM, 2005).