GEOPHYSICS, VOL. 62, NO. 4 (JULY-AUGUST 1997); P. 1045–1046

To Bayes or not to Bayes?

John A. Scales* and Roel Snieder‡

Editor's Note:

A goal that I have set for myself as Editor is to add some new features to our standard fare of erudite technical articles. There are always times when those who do science need to pause and reflect about the broader implications of what they do. It occurred to me that an occasional column by writers who want to probe into such matters might interest a large cross-section of our readers. Drs. John Scales and Roel Snieder of the Colorado School of Mines and Utrecht University have written a fascinating piece on a geophysical inversion methodology associated with the name of the famous eighteenth century mathematician and clergyman Thomas Bayes. I hope that you will enjoy their short essay. I would also like to know how you, the readers, feel about columns of this kind, and would welcome your feedback, preferably by e-mail to streitel@seg.org. I also encourage you to send me your specific comments about the present essay.

—Sven Treitel
Editor

The goal of geophysical inversion is to make quantitative inferences about the Earth from noisy, finite data. The limitations of noise and the inadequacy of the data mean that geophysical inversion problems are fundamentally problems of statistical inference. We do not invert data to find "models," as much as we might like to; we invert data to make inferences about models. There will usually be an infinity of models that fit the data. Thus we must look to probability theory for help.

There are two fundamentally different meanings of the term "probability" in common usage. If we toss a coin N times, where N is large, and see roughly $N/2$ heads, then we say the probability of getting a head in a given toss is about 50%. This interpretation of probability is therefore called "frequentist." On the other hand, you can't turn on the evening news without hearing a statement such as: "the probability of rain tomorrow is 50%." Since this statement does not refer to the repeated outcome of a random trial, it is not a frequentist use of the term probability. Rather, it conveys a statement of information. This is the Bayesian use of "probability." Both ideas seem quite natural, so it is perhaps unfortunate that the same term is used to describe them.

Bayesian inversion has in recent years gained a strong popularity in its application to geophysical inverse problems. The philosophy of this procedure is as follows. Suppose one knows something about a model before using the data. This knowledge is cast in a statistical form and is called the a priori model information. (A priori means *before* the data have been recorded; i.e., information that is independent of the data.) Suppose one then has a set of data, and that one also knows the statistical properties of the data (e.g., the data variance and covariance). Bayesian inversion provides a framework for combining the a priori model information with the information contained in the data to arrive at a more refined statistical distribution; that is, the a posteriori model distribution. The a posteriori distribution is what we know *after* we have assimilated the data and our prior information. The point of using the data is that the a posteriori model information hopefully constrains the model more tightly than the a priori model distribution.

The popularity of Bayesian inversion cannot be explained by the conceptual elegance of the method only. Instead, the popularity is to a large extent a result of the freedom that is taken in controlling the desired model properties through the specification of the a priori model statistics. In other words, the a priori knowledge of the model is often used as a knob to tune the properties of the final model produced.

However, the notion of a priori model statistics can in practice be somewhat shaky. As an example, consider a seismic survey. In such a situation one may have a fairly accurate idea of the ranges of seismic velocity and density that are realistic. However, the length scale of the velocity and density variation is to a large extent unknown. Sonic logs taken over increasingly smaller lengths indicate that the earth's properties might not have a characteristic length scale at all. The horizontal correlation length is a property that is almost impossible to assess

*Dept. of Geophysics and Center for Wave Phenomena, Colorado School of Mines, Golden Colorado 80401. E-mail: jscales@dix.mines.edu.
‡Dept. of Geophysics, Colorado School of Mines and Utrecht University, P.O. Box 80.021, 3508, Utrecht, The Netherlands. E-mail: snieder@geof.ruu.nl.

a priori, especially if one takes into account that seismic surveys often are carried out to detect the rapid horizontal variations in the seismic velocity that occur near structural features such as faults or salt domes. This implies that our a priori knowledge of the earth's interior is rather poor, especially where it concerns the a priori correlation length of the model.

So how can Bayesian inversion be so popular when our a priori knowledge is often so poor? The reason for this is that in practice one often uses the a priori model statistics to regularize the a posteriori solution. In other words, in a succession of different inversions one tunes the a priori model statistics in such a way that the retrieved model has agreeable features. Note that in such an approach the a priori model statistics are used a posteriori to tune the retrieved model!

But Bayesian statistics relies completely on the specification of a priori model statistics, i.e., on the knowledge that one has of the model *before one has used the recorded data*. (This specification is encapsulated in Bayes's theorem, which can be found in textbooks on probability.) The flexibility taken in using the a priori model statistics as a knob to tune properties of the retrieved model therefore is completely at odds with the philosophy of Bayesian inversion. This does not mean that there is anything wrong with Bayesian inversion, but it does imply that the reason for the popularity of Bayesian inversion within the earth sciences is inconsistent with the underlying philosophy. The quote "if I hadn't believed it I would not have seen it" certainly applies to this type of abuse of Bayes's theorem.

In practice it is extremely difficult to use Bayes's theorem to do realistic inverse problems and be honest. In doing so it is crucial to answer the following two questions:

1) Is the mathematical representation of the information justified by the observations? This applies both to the a priori model information (as we have seen), but also to the description of the data statistics.
2) To what extent does this information refine the inferences?

For Bayesians, the second question is the easiest to answer. If we have a probability that assimilates all the available information, then we can compute a variety of measures telling us how reliable our estimates are. Essentially, this means that we can define bounds for the model parameters, centered about a most-probable value, within which we are confident, to a certain degree, that the parameters must lie. The higher the confidence we specify, the bigger the set of possible parameter values. These are our "error bars."

It's the first question that causes problems. For realistic inverse problems one has a choice of three strategies. The first strategy is to apply Bayes's theorem and prescribe the a priori model statistics and data statistics "as honestly as possible." (Note how difficult it is to give a formal definition of this last statement.) In doing so one discovers that there are few objective criteria to define the a priori model and data statistics. For example, is it justified to take a statement such as: parameter x must lie between a and b and represent this statement as a uniform probability on the interval? No! There are infinitely many probability distributions consistent with this statement. To pick one is an overspecification of the information given. Even an apparently conservative approach such as taking the probability distribution that maximizes the entropy subject to the constraint that x lies in the interval may lead to pathologies in high-dimensional problems. This implies that it is extremely difficult to prescribe unambiguously the statistical properties of the data and the a priori model. One way out of this dilemma is to presume that "probability lies in the eye of the beholder." Another possibility is to derive priors based on a fundamental theory. We shall call this approach a pragmatic Bayesian strategy.

A second strategy, more purely Bayesian, is to attempt to model all quantitative information. (The distinction between data and information is fuzzy at best. The point is that much of the information that goes into the priors of the pragmatic Bayesians is ultimately "data" and therefore subject to modeling.) This means that the data misfit function can be generalized to include all such information. What's left over goes into the a priori distribution and should be as noninformative as possible. This limits the use of subjective priors, but still requires the estimation of the statistics of all the information, now regarded as data. But even noninformative priors are really informative inasmuch as they are often based on theories and observations. It is important to be able to show that the resulting inferences do not depend strongly on this prior. In addition, for practical purposes, a truly "noninformative" prior distribution is often not very useful since it may primarily reveal that our data are consistent with a shockingly large class of models.

A third strategy is to abandon Bayes altogether and use only deterministic prior information about models; density is positive, for instance. The inference problem is still statistical since random data uncertainties are taken into account. Essentially the idea is to look at the set of all models that fit the data. Then perform surgery on this set, cutting away those models that violate the deterministic criteria, e.g., have negative density. The result will be a (presumably smaller) set of models that fit the data and satisfy these a priori considerations. If we do this, however, we must accept that any particular model in this set, the least-squares model for instance, has no special significance, even though it may fit the data slightly better than the other models in the set. We have no way of saying that model x is more probable than model y, since in this frequentist approach we do not attempt to put a probability distribution on the models themselves.

A common complaint we hear is that people do not want to hear about ranges of models, error bars, or all this messy statistics. Perhaps the answer is to avoid speaking of "inverting data." Perhaps we should tell people that we are measuring the risks associated with various interpretations of the data. After all, interpreters do not really want to have a color plot of model parameters. They want to have answers to concrete questions. What are my chances, say, of penetrating a certain lithologic boundary if I drill to a certain depth? An answer of the following form would be highly valuable: With 95% confidence, the depth to this target is between a and b. If the difference between a and b is small, the interpreter will know that the risks are similarly small.

So the answer to the question posed in the title would seem to be: yes, but only when one can justify characterizing the prior information probabilistically, which, as we have indicated above, is a good deal harder than apparent at first glance.