A Tale of Two Chains: Experiences from a Collaboration Between Statisticians and Physicists

Robert H. Swendsen and Brian Diggs
Department of Physics, Carnegie-Mellon University, Pittsburgh, Pennsylvania

Christopher Genovese and Joseph B. Kadane
Department of Statistics, Carnegie-Mellon University, Pittsburgh, Pennsylvania

Shing-Te Li
Department of Physics, Carnegie-Mellon University, Pittsburgh, Pennsylvania

Abstract. It was the best of situations. It was the worst of situations. Markov Chain Monte Carlo techniques have played an important role in both physics and statistics, and yet, while the algorithms may be the same in these fields, the perspectives with which the techniques are used are strikingly different. With both statisticians and physicists in our group, we have learned a number of interesting lessons during our long collaboration, regarding both methodology and philosophy. In this paper, we relate some of our experiences in the context of a statistical problem in thermal physics. Perhaps these observations will help collaboration between these two fields reach a far, far better place.

1. Introduction

The views presented in this paper, and in the talk given at the conference, have evolved from an ongoing collaboration between a group of statisticians and physicists. The collaboration has been intellectually stimulating, occasionally confrontational, and very rewarding. As in many successful collaborations, it has gone off in unexpected directions, and revealed things about our individual assumptions that we had not previously thought to be at issue. In keeping with the intended, and successfully realized, interdisciplinary nature of this conference, we would like to share some of our experiences, along with the insights we have found while exploring paths that were not previously known to exist. The text of this paper was written after the conference. It will attempt to record some consequences of the rich discussion that took place in conjunction with both this talk and during the rest of the conference. Those of us who attended the conference (J.B.K. and R.H.S.) were very impressed by the wide-ranging and knowledgeable discussions that included such diverse physics topics as renormalization group theory and Anderson localization. At the end of this paper, we will make some comments about some possible future directions of research in oceanography, which might combine the extraordinary progress oceanographers have made in modeling the flow of ocean currents, with the viewpoints of renormalization-group studies. This was not the first interdisciplinary collaboration for the members of our group. Indeed, we had had extensive experience in other such collaborations, which prepared us for some of the features of our discussions. An important role that we have found is that whenever somebody from another field seems to have taken leave of their senses and is saying things that cannot possibly be true, pay attention! What you are hearing is often something that is both true and obvious to people in another discipline. It is the most clearly distinguishing earmark of an impending breakthrough.

2. The Language of Collaborations

The first hurdle any collaboration has to get over is the illusion that we speak a common language. If two languages are obviously different, like French and Japanese, an immediate effort is made to learn the language of your colleagues, or at least find an accurate method of translation. However, the languages of different scientific fields use many words in common, often creating the erroneous impression that we are all talking about the same thing. Some differences in terminology are immediately recognizable. In performing computer simulations, physicists talk of the "equilibration" of the system, in analogy to the time required in experiments to make sure that different components of the system are in thermal equilibrium with each other. Statisticians use the more colorful term, "burn in," to describe the same aspect of the simulation. Other terms can be more difficult. The physicist's term for an efficient approach to simulation using random numbers is the "Monte Carlo" method [*Binder*, 1992]. The statistician's term for exactly the same method is "Markov Chain Monte Carlo," [*Gilks et al.*, 1996] because an essential feature of the method is setting up a Markov chain to preferentially sample the important states of the system. This would not be too bad, if it weren't for the fact that statisticians use the term "Monte Carlo" by itself to mean a kind of random sampling that physicist would refer to as an infinite-temperature simulation. The use of the term "Monte Carlo" does not, therefore, always produce a flag that something is wrong, since the other meaning might be regarded as possible in the context of some discussions. This sort of gap can exist even within a discipline, and one of us (R.H.S.) has encountered dramatic differences in vocabulary even between different fields of physics. In discussing Monte Carlo simulations of lattice gauge models, he discovered that the terms "action" (particle physics) and "Hamiltonian" (statistical mechanics) refer to the same concept. Even more confusing is the term "mass" in particle physics, which corresponds to an inverse correlation length in statistical mechanics. However, the worst confusions were caused by the use of "strong" and "weak" coupling, which have reversed meanings in the two fields.

3. From Vocabulary to Patterns of Thought

Some of the differences in terminology reflect different patterns of thought. Encountering these differences can be extremely frustrating at first, but very fruitful later on in the collaboration in opening up the range of possibilities for attacking problems. One example of this is the apparently innocuous word, "model," which we discussed at the conference. In physics, "model" means a simplified mathematical representation of a real system. An example that formed a focal point of our collaboration is the Ising model of magnetism, which represents molecular magnetic moments with simple variables located on the sites of a lattice, and point either up or down in response to a magnetic field and interactions with each other. In statistics, "model" means an assumed mathematical relationship between data and a set of parameters, plus knowledge (or beliefs) about the values of the parameters. The discussion that this distinction raised during the conference clearly illustrated the kinds of difficulties it can cause. There was an immediate focus on which definition was "correct," with different participants presenting different points of view. The essential point is that both definitions are correct. Definitions are arbitrary, and they are defined by the members of a group in a way that they have found by experience to be useful. Another example of a key word with different meanings is "data." In statistics, "data" refers to the result of a measurement of some quantity that describes the real world. In physics, the word is used in a much more general sense, which includes numbers obtained from computer simulations. The physics definition probably arose from the early treatment of simulations as experiments performed on a computer. Different definitions can reflect different ways of approaching the analysis of a problem, and it is precisely the interactions between methods from one field and problems from another field that produce the most powerful collaborative results.

4. What We Have Learned From the Collaboration

For the physicists in the group, the most dramatic change brought about by the collaboration has been the introduction to Bayesian methods and the accompanying patterns of thought. The distinction between "frequentist" and "Bayesian" views of probability is familiar to most people at the conference. Put briefly, the frequentist regards probability as the ratio of successes to the total number of trials in a long series of experiments. The Bayesian view is that probability is an expression of a degree of belief based on evidence. For a statistician looking at a physics department, it is startling to realize that almost all physics books dealing with probability and statistics give the frequentist view as the only valid possibility [*Reif*, 1965; *Bevington and Robinson*, 1992]. The Bayesian approach is rarely discussed, thus depriving physicists of a very useful set of tools and modes of thought for dealing with the problems they encounter. Learning to use Bayesian methods has proved to be extremely valuable. In particular, our understanding of what you are really assuming when you truncate a series has changed drastically as a result of this collaboration, giving rise to an entirely new way of analyzing series expansions that is beyond the scope of this paper. For the statisticians in the group, there were a number of approaches to performing and analyzing (Markov Chain) Monte Carlo simulations that have proved to be valuable in other contexts. Some of these are discussed explicitly below.

5. (Markov Chain) Monte Carlo

The main theme of our collaboration is not restricted to the explicit properties of Ising models, which are admittedly of limited interest to oceanographers, but deals with understanding the properties of general physical systems. The unifying methodology is the Bayesian approach and the use of Monte Carlo (or Markov Chain Monte Carlo) computer simulations to study the model [statistics definition] under consideration. Because the use of MC (MCMC) in such analyses is increasingly becoming a standard tool, we hope our results will have a wide applicability. The essential feature of (Markov Chain) Monte Carlo simulations is the creation of stochastic process for creating sequences of states of the system under consideration that leads to states being selected according to their importance. In physics, this importance might be their thermal probability; in statistics, this might be the Bayesian posterior distribution. In all cases, an equation of detailed balance is constructed as cleverly as possible, and sequences of states are realized with the use of pseudo-random numbers. The method is attractive because of its relative simplicity, efficiency, and applicability to a wide range of problems. We would like to comment on two aspects of these simulations that occur in almost all applications. In fact, the problems we will address also occur in simulations using methods that might seem to be based on entirely different principles. We believe that they are particularly applicable to computer simulations of the flow patterns in the ocean. Which is better: one long simulation, or many short simulations? The usual answers to this question depend on the field of application [*Geyer,* 1992; *Gelman and Rubin,* 1992 a,b]. This partly reflects the different natures of the probability distributions encountered in different sorts of applications. As people accumulate experience with one class of problems, they develop standard approaches that work effectively, but which might break down in more complex situations. Although it is often the case that one long simulation is superior to many short ones, this can be very dangerous when the global properties of the system are not well known (which, of course, is the most useful case). It is quite possible for the probability distribution under investigation to have multiple maxima, each of which is plausible, but different in important ways. This phenomenon is known in physics as the problem of metastable states. If a system is in a metastable state, it's parameters are near a local maximum in the probability distribution, but not near the global maximum that represents the most probable behavior. Metastability is not a defect in the mathematical representation, but usually reflects an important aspect of reality. Shifting patterns in ocean currents provide examples of metastable states. Ice ages are metastable states in the world's climate, in which different states are thought to be roughly equally probable. In performing a computer simulation, it is perfectly possible for a correct program of a valid model to produce a result that differs from observations if a metastable state is encountered. It is very difficult, if not impossible, to detect metastability from a single computer run. However a series of runs made from different starting points creates the possibility of detecting metastable states by the different predictions of different runs. Since encountering and failing to recognize a metastable state can invalidate the results of a simulation, it is comforting to know that different runs give the same result. A technical warning: If the initial state of a simulation is varied only by using a new set of random numbers, it is possible to encounter the same metastable state in every run. This can give a false impression of security when the results are incorrect. It is important to use your knowledge of the system to construct starting states that differ in significant ways. An illustration from physics would be starting from a random configuration of spins or particles. This would correspond to starting at infinite temperature, regardless of the particular set of random numbers. Starting in an ordered or crystalline state can, in some cases, produce very different behavior. How long do you have to run a simulation? A full analysis of this question depends on the concept of a correlation time, t_{corr}, such that the correlations between states decay as $\exp[-t/ t_{corr}]$. The burn-in or equilibration time, t_{eq}, must be much greater than t_{corr}. If N runs with different starting configurations are simulated for a total time t_{TOTAL}, the error will be proportional to *Müller-Krumbhaar and Binder* [1973].

$$[(1 + 2\, t_{corr}) / (t_{TOTAL} - \mathrm{N}\, t_{eq})]{+1/2}.$$

Clearly, it is important to have at least a rough idea of the magnitude of the correlation time. If that information can be obtained, it is fairly simple to determine how long to run a simulation to obtain the desired accuracy. Note that if $t_{TOTAL} >> \mathrm{N}\, t_{eq}$, there is little loss of accuracy from making several runs, so that testing for metastability is not very expensive.

6. Some Observations and Speculations About Oceanography

A final observation that we made during the very interesting discussions at the conference is that there is a very strong similarity between the properties of the oceanic flow patterns seen in computer simulations and the properties of magnets and fluids near a critical point. A "critical point," such as the temperature at which a magnet loses its magnetization, is characterized by (a) fluctuations on all length scales, from atomic through the maximum dimensions of the system, (b) interactions between fluctuations on different length scales, so that the large-scale properties of the system depend sensitively on the small-scale properties, and (c) large fluctuations at the longest length scales, possibly taking the system from one apparently metastable state to another. These properties are at least approximately true for models of the ocean, the atmosphere, and the long-term prediction of the climate. The sensitivity of the properties of the system to small-scale fluctuations is reflected in the importance of the grid size, which limits the representation of small-scale fluctuations. The chief value of this analogy lies in the possibility of dealing with the problem of grid size by using the methods of renormalization-group theory to calculate effective interactions between the variables at the grid points. It is possible, in principle, to model (physics definition) the relationship between different grid points so as to correctly reproduce the large-scale patterns without introducing more fitting parameters. Although it is impossible to anticipate the technical problems that might arise, we believe that this line of research might be extremely fruitful.

7. Conclusion

We believe that this overview of an interdisciplinary collaboration is characteristic of such collaborations. It has ranged over a wide variety of topics that might seem disconnected at first, but are really closely tied together in ways that enhance the understanding both fields more deeply.

References

Bevington, P. R., and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw Hill, New York, 1992.

Binder, K., *Topics in Applied Physics*, Springer-Verlag, Berlin, 1992.

Gelman, A., and D. B. Rubin, A single series from the Gibbs sampler provides a false sense of security, in *Bayesian Statistics 4*, edited by J. M. Bernardo, J. Berger, A. P. David, and A. F. M. Smith, pp. 625-631, Oxford University Press, 1992a.

Gelman, A., and D. B. Rubin, Inference from iterative simulation using multiple sequences, *Statist. Sci., 7*, 457-472, 1992b.

Geyer, C. J., Practical Markov chain Monte Carlo, *Statist. Sci., 7*, 473-511, 1992.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London, 1996.

Müller-Krumbhaar, H., and K. J. Binder, Dynamic properties of the Monte Carlo method in statistical mechanics, *Stat. Phys. 8, 1*, 1973.

Reif, F., *Fundamentals of Statistical and Thermal Physics*, McGraw Hill, New York, 1965.