

How to Measure What We Don't Know

September 10 2009, By Lisa Zyga

(PhysOrg.com) -- How do we discover new things? For scientists, observation and measurement are the main ways to extract information from Nature. Based on observations, scientists build models that, in turn, are used to make predictions about the future or the past. To the extent that the predictions are successful, scientists conclude that their models capture Nature's organization. However, Nature does not reveal secrets easily - there is no way for observers to learn everything about a process, so some information always remains hidden from view; other kinds of information are present, but difficult to extract. In a recent study, researchers have investigated how to measure the degree of hidden information in a process (its "crypticity") and, along the way, solved several puzzles involved in extracting, storing, and communicating information.

In their study, James Crutchfield, Physics Professor at the University of California at Davis, and graduate students Christopher Ellison and John Mahoney, have developed the analogy of scientists as cryptologists who are trying to glean hidden information from Nature. As they explain, "Nature speaks for herself only through the data she willingly gives up." To build good models, scientists must use the correct "codebook" in order to decrypt the information hidden in observations and so decode the structure embedded in Nature's processes.

In their recent work, the researchers adopt a thorough-going informational view: All of Nature is a communication channel that transmits the past to the future by storing information in the present. The information that the past and future share can be quantified using the

“excess entropy” - the mutual information between the past and the future.

Since the present mediates between the past and future, it is natural to think that the excess entropy must somehow be stored in the present, the researchers explain. And while this is true, the researchers showed that, somewhat surprisingly, the present typically contains much more information than just the excess entropy. The information stored in the present is known as the “statistical complexity.” The more information Nature must store to turn her noble gears, the more structured her behavior.

The information that manages to go unaccounted for - the difference between the stored information (statistical complexity) and the observed information (excess entropy) - is the “crypticity”. It captures a new and under-appreciated complexity of a process, something that goes above and beyond what is directly measured in observations. At a more general level, the researchers provide an explicit way to understand the difference between simply making predictions from data versus modeling the process’s underlying structure.

“The results are at the crossroads of several research threads, from causal inference to new forms of computing,” Crutchfield told *PhysOrg.com*. “But here are a couple of things we highlight: One can look at all of nature as a communication channel: Nature communicates the past to the future, by storing information in the present. In addition, information about how a system is structured can be available in observations, but very hard to extract. Crypticity measures the degree of that difficulty. Even in equilibrium there are temporal asymmetries.”

Although excess entropy, statistical complexity and crypticity are straightforward to define, their direct calculation has been a long-standing puzzle. Crutchfield, Ellison, and Mahoney developed a novel

approach to its solution. The process, interpreted as a communication channel, is scanned in both the forward and reverse time directions to create models for prediction and retrodiction. By analyzing the relationship between predicting and retrodicting, they were able to uncover not only the external, time-symmetric information (excess entropy), but also the internal, asymmetric information (statistical complexity and crypticity). By looking inside Nature's communication channel, they discovered a rather non-intuitive asymmetry: Even processes in equilibrium commonly harbor temporally asymmetric structures.

“The basic idea is that a process can appear to not transmit much information from its past to its future, but still require a large amount of hardware to keep the internal machine going,” Crutchfield said. “For example, imagine that you have two coins: Coin A is a fair coin and Coin B is slightly biased. Now the output of this process is a series of heads and tails. That's all the observer gets to see. The observer doesn't know when A is used or B is used. To an observer this process is very close to a fair coin - the heads and tails from B just don't differ much in their statistics from the heads and tails from A. So, the observed process has little mutual information (the heads and tails are pretty much independent of the past). That is, the process has very low excess entropy. Nonetheless, there is one bit of internal stored information: Which coin, A or B, is flipped at each step? You can take this example to an extreme where you have hundreds of internal coins, all slightly biased, all slightly different in their bias, and therefore distinct coins. The large number of coins gives you an arbitrarily large statistical complexity. But the small biases mean the excess [entropy](#) is as close to zero as you like.”

These fundamental results should impact research across a wide range of disciplines, from statistical modeling to novel forms of computing. As the researchers explain, when a process contains hidden information, the

process cannot be directly represented using only raw measurement data. Rather, a model must be build to account for the degree of hidden information that is encrypted within the process's observed behavior. Otherwise, analyzing a process only in terms of observed information overlooks the process's structure, making it appear more random than it actually is.

“In statistical modeling, if you ignore a process's crypticity, you will conclude that nature is more random and less structured than she really is,” Crutchfield said. “We suspect that this general principle will be seen (or is even operating) in many scientific domains, from biosequence analysis to dark energy modeling.”

Intel, which partially funded this research, also has an interest in using the results to improve network performance. For many years, Intel has funded research on complex systems through the Network Dynamics Program, which Crutchfield runs.

“Intel's original interest, 10-plus years ago, was to stimulate research on the structure and dynamics of networks,” Crutchfield said. “This was an extremely successful program, in fact stimulating much progress in the early years that has now blossomed into the field of network science. The present work adds to that growing body of understanding of how complex systems are organized and how we should model and characterize them. In particular, crypticity and causal irreversibility suggest new metrics for network performance.”

More information: James P. Crutchfield, Christopher J. Ellison, and John R. Mahoney. “Time's Barbed Arrow: Irreversibility, Crypticity, and Stored [Information](#).” *Physical Review Letters* 103, 094101 (2009).

Copyright 2009 PhysOrg.com.

All rights reserved. This material may not be published, broadcast,

rewritten or redistributed in whole or part without the express written permission of PhysOrg.com.

Citation: How to Measure What We Don't Know (2009, September 10) retrieved 19 April 2024 from <https://phys.org/news/2009-09-dont.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.