

About the Cover

Data MINEing

April is Math Awareness Month, and this year the theme is data mining. The cover illustrates a simplified variant of the impressive tool MINE described in a recent *Science* article by David and Yakir Reshef and several others, “Detecting novel associations in large data sets”, in volume 334 from last December).

“Data mining” is the term coined to describe what one does in order to extract valuable information from the huge amounts of (likely noisy) data that modern computers make available. Among the information one might want to extract are the relations that hold among various specific variables recorded in statistics, for example car accident rate and age. But one might also be given data on a large number of variables, and wonder without much prior information which of them are in fact related. This is what MINE helps with—it scans data for whatever pairs, possibly all, that you might think of interest, and produces for each pair a number—in fact a whole matrix of numbers—that measures the strength of relationship. The basic technique is to fit grids of various sizes to a set of pairs (x,y) in the plane in order to maximize the *mutual information coefficient* of data in the grid. Roughly speaking, the mutual information coefficient (*mic*), based on Shannon’s information theory and first defined by the astronomer E. H. Linfoot in 1957, tells how much information about one variable is implied by the other. The cover shows a collection of sample small grids on artificial data.

The problem that Reshef, Reshef, et al. attack looks at first almost impossible. After all, the number of ways to partition a large planar set is (so to speak) astronomical. But because of the additive properties of entropy, a sub-grid of an optimal grid must also be optimal, and this suggests an approach by dynamic programming that turns out to make the task feasible.

One of the authors of the article has described some features of MINE, as well as the process of publishing in *Science*, on his website:

<http://mybiasedcoin.blogspot.com/2011/12/mic-and-mine-short-description.html>

—Bill Casselman
Graphics Editor

(notices-covers@ams.org)