

STANFORD UNIVERSITY  
DEPARTMENT OF STATISTICS  
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, May 29, 2001  
Sequoia Hall Rm. 200  
(Cookies at 3:45 in 1st Floor Lounge)

*David Donoho*  
*Department of Statistics*  
*Stanford University*  
*Stanford, CA 94305*

**The Kolmogorov Sampler**

Recently a considerable body of work in model selection has focused on estimation by maximizing a penalized likelihood, where the penalty is an intuitive measure of complexity, such as no.(terms in model). This criterion, it is argued, favors the “simplest” model which explain the data adequately.

Suppose instead we measure complexity of the bitstring  $y$  by the length of shortest computer program that computes it. This notion is often called Kolmogorov complexity, and has captivated many as the most satisfying notion of complexity.

So, given noisy data  $x$ , suppose we look for the lowest complexity  $y$  that matches the observed  $x$  to within the noise level according to the natural goodness of fit measure associated with the noise distribution.

This can be viewed as the ultimate form of complexity-penalized likelihood estimation, in which the complexity penalty is measured in its ultimately most satisfying way. We might call this the Kolmogorov estimator. We remark that it is a deterministic function of the input data.

We consider performance of this estimator in a few interesting continuous and discrete settings where it is applied to noisy data generated by a Bayesian prior for the many unknown estimands. We show that, in these settings anyway, the estimator quite generally and quite precisely acts as a sampler from the underlying posterior distribution from whatever prior happens to be the case – this, despite the fact that no prior knowledge about the object to be recovered has been input to the procedure, but only some information about the noise distribution. In other words, the minimum complexity criterion automatically ‘figures’ out what prior distribution is holding, and generates a sample from

the corresponding posterior. So we had better call this the ‘Kolmogorov Sampler’, even though it is a deterministic function of the data.

It follows, for example, that in estimating a high-dimensional mean vector in Gaussian noise, for squared-error loss and under a variety of assumptions about the means e.g.(a) iid means from an unknown prior or (b) means following an unknown stationary process, the Kolmogorov Sampler achieves asymptotically twice the Bayes risk. So it is quite impressively adaptive, although only 50

Why is this true? The argument is based on a chain of relationships between Kolmogorov Complexity, Shannon Rate Distortion Theory, and Bayes Decision Theory which I will explain. Some of these do not seem to be very well known, even to experts, and perhaps they should be better known.