

## Re: Induction of statistical models

**Source:** <http://sci.tech-archive.net/Archive/sci.stat.math/2004-12/0799.html>

---

**From:** Herman Rubin ([hruhin\\_at\\_odds.stat.purdue.edu](mailto:hruhin_at_odds.stat.purdue.edu))

**Date:** 12/23/04

Date: 23 Dec 2004 11:45:51 -0500

In article <cqeqrI\$j5k\$1@planja.arnes.si>, Aleks Jakulin <a\_jakulin@hotmail.com> wrote:

>Ross asked:

>> After George's comment, I'd like to rephrase my question to be: "Can  
>> anyone recommend papers, books, or other resources that describe  
>> building, either from the point of view of somebody training to be a  
>> statistician, or from the point of view of building systems that can  
>> automate model building?"

>This question reminds me of someone asking for good books on  
>mathematics. :) Model building is a tremendously broad topic, with  
>several different schools of thought.

>We've had a few discussions recently, though:

>1. Why Occam's razor "If there are several hypotheses equally  
>consistent with the data, pick the simplest one". Malcolm Forster's  
>writing on the topic is quite lucid:

><http://philosophy.wisc.edu/forster/> In a similar vein, there is a  
>whole book with "model selection" in the title. I've just read it, and  
>it's quite

>reasonable:

>*Model Selection and Multi-Model Inference*

><http://www.springeronline.com/sgw/cda/frontpage/0,11855,5-10129-22-2009034-0,00.html>

>2. However, Bayesians disagree with Occam's razor, and with model  
>selection in general:

>[http://www.stat.columbia.edu/~cook/movabletype/archives/2004/12/against\\_parsimo.html](http://www.stat.columbia.edu/~cook/movabletype/archives/2004/12/against_parsimo.html)

There is no basic disagreement between Bayesian inference and Occam's razor, if the Bayesian inference is done properly, not rashly. One will never get the true model, nor could one use the true model if by some miracle it can be found, so the real question is, what somewhat wrong model minimizes the combined aspects of the loss? The behavioral Bayes approach does not look at the probability of the right action, nor does it even require prior probability in the usual sense, but minimization of the

sci.stat.math: Re: Induction of statistical models

Bayes risk, and even this can only be approximated in practice. Some of the aspects of risk are the complexity of the model, the error in predicting from the model, computational costs, whether the model can improve understanding, etc.

>3. *On the Epicurean principle "It would be unscientific to choose an arbitrary hypothesis if several are consistent with the data", and a possible synthesis with Occam's razor:*  
>[http://www.stat.columbia.edu/~cook/movabletype/archives/2004/12/wacky\\_computer\\_1.html](http://www.stat.columbia.edu/~cook/movabletype/archives/2004/12/wacky_computer_1.html)

>--

--

This address is for information only. I do not claim that these views are those of the Statistics Department or of Purdue University.  
Herman Rubin, Department of Statistics, Purdue University  
hrubin@stat.purdue.edu Phone: (765)494-6054 FAX: (765)494-0558