

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENTAL SEMINAR

4:15 p.m., Thursday, June 29, 2000
Sequoia Hall Rm. 200
(Cookies at 3:45 in 1st Floor Lounge)

Peter Grunwald
EURANDOM
The Netherlands

SAFE STATISTICS

Statistical analysis of data more often than not results in inference of a model that is “wrong yet useful”: it is wrong in that it is really a gross simplification of the process actually underlying the data; it is useful in that decisions and predictions about future data taken on the basis of the model are quite successful. Examples of this phenomenon abound: we assume (conditional) independencies between random variables which really are highly dependent (i.e. in speech recognition based on hidden Markov models); we assume linear models for non-linear phenomena; we model noise by a Gaussian also in cases where noise is really not Gaussian at all – and we often get away with this.

Here we give a novel explanation of this phenomenon. We show that there exist models (the simplest examples are the i.i.d exponential families) that can be ‘safely’ used for inference even if the ‘truth’ is not even close to any of the distributions under consideration. We outline a general theory that allows us to determine under exactly what circumstances it is useful or even advisable to use such overly simple but ‘safe’ models for the data at hand. We discuss relations to other work on ‘robust’ inference under misspecification.