

## REMEMBERING LEO BREIMAN

BY RICHARD A. OLSHEN

*Stanford University*

I published an interview of Leo Breiman in *Statistical Science* [Olshen (2001)], and also the solution to a problem concerning almost sure convergence of binary tree-structured estimators in regression [Olshen (2007)]. The former summarized much of my thinking about Leo up to five years before his death. I discussed the latter with Leo and dedicated that paper to his memory. Therefore, this note is on other topics. In preparing it I am reminded how much I miss this man of so many talents and interests. I miss him not because I always agreed with him, but instead because his comments about statistics in particular and life in general always elicited my substantial reflection.

Technical comments here are in part my responses to Leo's 2001 paper in *Statistical Science* [Breiman (2001)]. The paper is interesting and provocative, but it demonstrates an attitude that seemed somewhat unfortunate in 2001 when it was published and remained so in 2005 when Leo died. It is even less fortunate today. D. R. Cox may have stated the obvious when he noted in his discussion [Breiman (2001), page 216] that, "Like all good caricatures, it contains enough truth and exposes enough weaknesses to be thought-provoking."

In his discussion of the paper, Bradley Efron states (page 219) that, "Prediction is certainly an interesting subject. Leo's paper overstates both its role and our profession's lack of interest in it... the whole point of science is to open up black boxes, understand their insides, and build better boxes for the purposes of mankind... we can hope that the present paper was written more as an advocacy device than as the confessions of a born-again black boxist."

For years I have preferred Cox's approach [Breiman (2001), page 216]. "Professor Breiman takes data as his starting point. I would prefer to start with an issue, a question or a scientific hypothesis." Also, I believe strongly that crisp mathematical formulations of statistical problems can clarify rather than obscure them; likely, if pressed Leo would have agreed. The paper and Bruce Hoadley's discussion of it focus on the importance of predictors. A predictor might be "important" if it predicts whatever outcome is in question accurately by itself. Alternatively, it might be called "important" if the performance of other predictors is harmed by its absence. More generally, a variable might be deemed important if it is approximately mutually predictable with a set of predictors, and the entire set is important or not by either criterion. These notions permit easy expression in mathematical