Annual Meeting of the Montana Chapter of the American Statistical Association

September 18, 2007

Montana Tech, Butte, MT

Scientific Session Program

8:45-9:30AM	COFFEE AND DOUGHNUTS
9:30AM	*Welcoming Remark* Solomon W. Harrar (Vice-president of the Chapter)
9:35-10:25	*Keynote Address* ***Occam's Razor and Statistical Model Selection*** Joseph E. Cavanaugh
10:25-10:50	***A Connection Between k-Nearest Neighbor Learners and Bagged Decision*** Joran Elias
10:50-11:15	***The History of Sir Gilbert Walker and His contributions to both Statistics and El Nino.*** Cindy Scavarda
11:15-11:40	***Misclassification Error Effects on Modeling Animal Habitat associations*** William R. Gould (presenter) and Zhiming Ni
11:40-1:10	LUNCH BREAK (No host lunch)
1:10-1:35PM	***Componentwise Variance Dispersion Graphs for Mixture Experiments*** John Borkowski
1:35-2:00	***Exact Bagging of k-Nearest Neighbor Learners*** Brian Steele
2:05	BUSINESS SESSION STARTS

Occam’s Razor and Statistical Model Selection

Joseph E. Cavanaugh
Department of Biostatistics
The University of Iowa

Occam's Razor is a pervasive philosophical principle that has been reflected in the writings of many renowned scholars, including Isaac Newton, Leonardo da Vinci, and Albert Einstein. The principle is often expressed as follows: given two or more competing explanations for a phenomenon, none of which can be discounted, the simplest explanation is to be preferred. Translated to statistical modeling, Occam’s Razor is often referred to as the Law of Parsimony, which claims that no more causes should be assumed than those that will account for the effect. In practice, models with a simplistic structure are often subjectively favored over models with a complex structure, since the former are more easily interpreted. However, from a statistical perspective, do simple models hold an advantage over complex models?

The statistical advantage of favoring models that comply with the Law of Parsimony is an improvement in inferential accuracy; in particular, in predictive accuracy. We show that the framework for discrepancy-based model selection criteria provides a paradigm for choosing a model that adheres to Occam’s Razor. We outline the development of this framework. In formulating a criterion, we discuss the importance of estimating a predictive measure known as the expected optimism, which leads to the criterion’s penalty term. The Akaike information criterion and its corrected variants are featured as examples.

We argue that model selection criteria may provide a more appropriate basis for model determination than hypothesis testing. To illustrate the utility of selection criteria, we consider applications based on an Australian coronary heart disease study and the Iowa Fluoride Study.

A Connection Between k-Nearest Neighbor Learners and Bagged Decision Trees

Joran Elias
Department of Mathematical Sciences
The University of Montana

We will present a brief introduction to some basic concepts in statistical learning theory (decision trees, bagging and k-nearest neighbor functions) and use these concepts to draw a connection between two seemingly disparate types of learners. We will also briefly discuss some difficulties that arise in attempting to exploit this connection and some possible avenues of exploration it may open up.

Misclassification Error Effects on Modeling Animal Habitat Associations.

William R. Gould and Zhiming Ni
Experimental Statistics Program
New Mexico State University

The linkage between habitats and perceived animal presence/absence has become a popular means of measuring habitat quality. Habitat use is compared to 'availability' and selection of habitats is inferred, often by using logistic regression modeling. One of the key assumptions is that animal presence and absence is measured with certainty. We investigated the effect of misclassifying animal presence and absence on parameter estimates of logistic regression models. Bias in parameter estimates was found to be substantial (over 50% relative bias) in most cases.

Researchers should survey multiple occasions, possibly using different survey methods, before assuming nondetection is equivalent to absence.

The History of Sir Gilbert Walker and His Contributions to both Statistics and El Nino.

Cindy Scavarda
The University of Montana

What do the Yule-Walker equations and El Nino have in common? The answer is Sir Gilbert Walker (1868 – 1958). His research focusing on monsoon prediction led to an explanation for the El Nino-Southern Oscillation as well as the development of the Yule-Walker equations. The discussion will largely be based on the paper by Richard Katz : *Sir Gilbert Walker and a Connection between El Nino and Statistics.* (Statistical Science, 2002). We will explore Walker's career, his contributions to statistics, and his research on the dynamics of El Nino.

Componentwise Variance Dispersion Graphs for Mixture Experiments

John Borkowski
Montana State University

Consider a q-component mixture experiment with constraints defined by lower and upper bounds for each component proportion. Whether the design region is a simplex or some irregularly shaped subregion, the problem of interest is the extension of variance dispersion graphs (VDGs) to experiments involving mixtures. As an alternative to studying the prediction variance on shrunken polyhedral spaces (Piepel and Anderson 1993) or to the prediction variance trace (PVT) plots involving Cox directions (Vining, Cornell, and Myers 1993), a new graphical tool, called the componentwise variance dispersion graph (CVDG), will be proposed. A CVDG contains componentwise plots of the variability throughout the entire polyhedral design space. Examples of CVDGs will be presented.

Exact Bagging of k-Nearest Neighbor Learners

Brian Steele
The University of Montana

Bootstrap aggregation, or bagging, is a method of reducing the prediction error of a statistical learner. In the context of the prediction problem, the goal of bagging is to construct a new learner which is the expectation of the original over the empirical or sample distribution function. In nearly all cases, the expectation cannot be computed analytically, and bootstrap sampling is used to produce an approximation. The k-nearest neighbor learners are exceptions to this generalization, and exact bagging is very easy. In addition to computational savings, there are interesting opportunities to study the bagging properties of k-nearest neighbor learners and to develop new exact and partially exact bagging learners.