AbstractID: 6006 Title: Bioinformatics, the Multiple-Biomarker Classifier Problem, Complexity, and Uncertainty

The most celebrated landmark of modern bioinformatics has been the sequencing of the human genome. Early in the project it was commonly believed that humans have about 100,000 genes and as the project neared completion the estimates came down into the neighborhood of 25,000-30,000. Hidden Markov Models (HMM) are used to carry out statistical parsing of the "linguistics" of such bioinformation. Such massively complex analysis has been facilitated by modern developments in massively complex hardware and software--but such analysis is naturally accompanied by great uncertainties. At a lower level of complexity are the algorithms used for computer-aided diagnosis in medical imaging, and at an intermediate level are the tools under current development for fusing multiple biomarkers--for example, from a large number of spectral lines in mass spectroscopy of protein fragments in blood samples and other mutliplex data from protein and gene microarrays. This talk will review the uncertainties in measured performance of such diagnostic tests as a function of the sample sizes available for training and testing as well as the dependence on the number of fused biomarkers and the complexity of the associated statistical learning algorithm. A strategy for designing large trials based on pilot studies will be outlined.

Educational Objectives:

1. Understand the multiple-biomarker classifier problem
2. Understand the uncertainties in its performance assessment due to finite training and testing
3. Understand the dependence on number of biomarkers and complexity of statistical learning algorithm