

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, October 9, 2007
Sequoia Hall Room 200
(Cookies at 3:45 in 1st Floor Lounge)

Brad Efron
Department of Statistics
Stanford University

Are a Set of Microarrays Independent of Each Other?

A typical microarray study might involve two groups of subjects, Controls and Treatments. Each subject provides material for his or her individual microarray, reporting some large number "m" of genetic expressions at the same time. In my main example, there are 63 subjects, 44 Controls and 19 Treatments, with $m = 20426$ genes, providing a 20426×63 data matrix X . We expect the rows of X to be correlated with each other since genes act in concert. The columns, that is the microarrays, are usually assumed to be independent. If, for example, we form two-sample t -statistics for each gene's data, the standard Student's t null hypothesis with $n-2$ degrees of freedom requires independence, as do familiar techniques such as cross-validation and permutation testing. This talk concerns testing a matrix X for column-wise independence when the rows may be highly correlated. The effect of row-wise correlation is to greatly reduce the power of standard tests. In my microarray example, row-wise correlation will be shown to reduce the effective sample size from 20426 to 17.