

Empirical Bayes Approach to Controlling Familywise Error: an application to HIV Resistance Data

Rhoderick N. Machekano
Division of Biostatistics
University of California, Berkeley

Abstract

New technologies such as micro-arrays have revolutionized the genetic study of human disease, generating massive datasets. Statistical challenges arise in identifying meaningful patterns and structures from these high dimensional genomic datasets. An example of such data comes from the sequencing of HIV genes targeted by antiretroviral drugs. When the HIV virus is under antiretroviral drug pressure, unfavorable mutations of the genes often lead to greatly increased resistance of the virus to the drugs, even to those drugs they are not exposed to. Identification of mutation combinations and their correlation to drug resistance is critical in guiding efficient prescription of HIV drugs. The identification of a subset of codons associated with drug resistance from a set of several hundreds of codons presents a multiple testing problem. We propose two multiple testing procedures (MTPs) controlling familywise error rate: an augmented-Bonferroni method and an Empirical Bayes procedure. Using simulations, we demonstrate that our proposed MTPs are less conservative and more powerful than the traditional methods such as Bonferroni and Holms procedures. We apply the methods to HIV resistance data where we wish to identify mutations in the protease gene associated with Amprenavir resistance.