

Why the Balding-Nichols Model is More Accurate In Calculating Genotype Frequency

Miguel Guardado and Rori Rohlfs,PHD

Department of Biology, San Francisco State University



Background

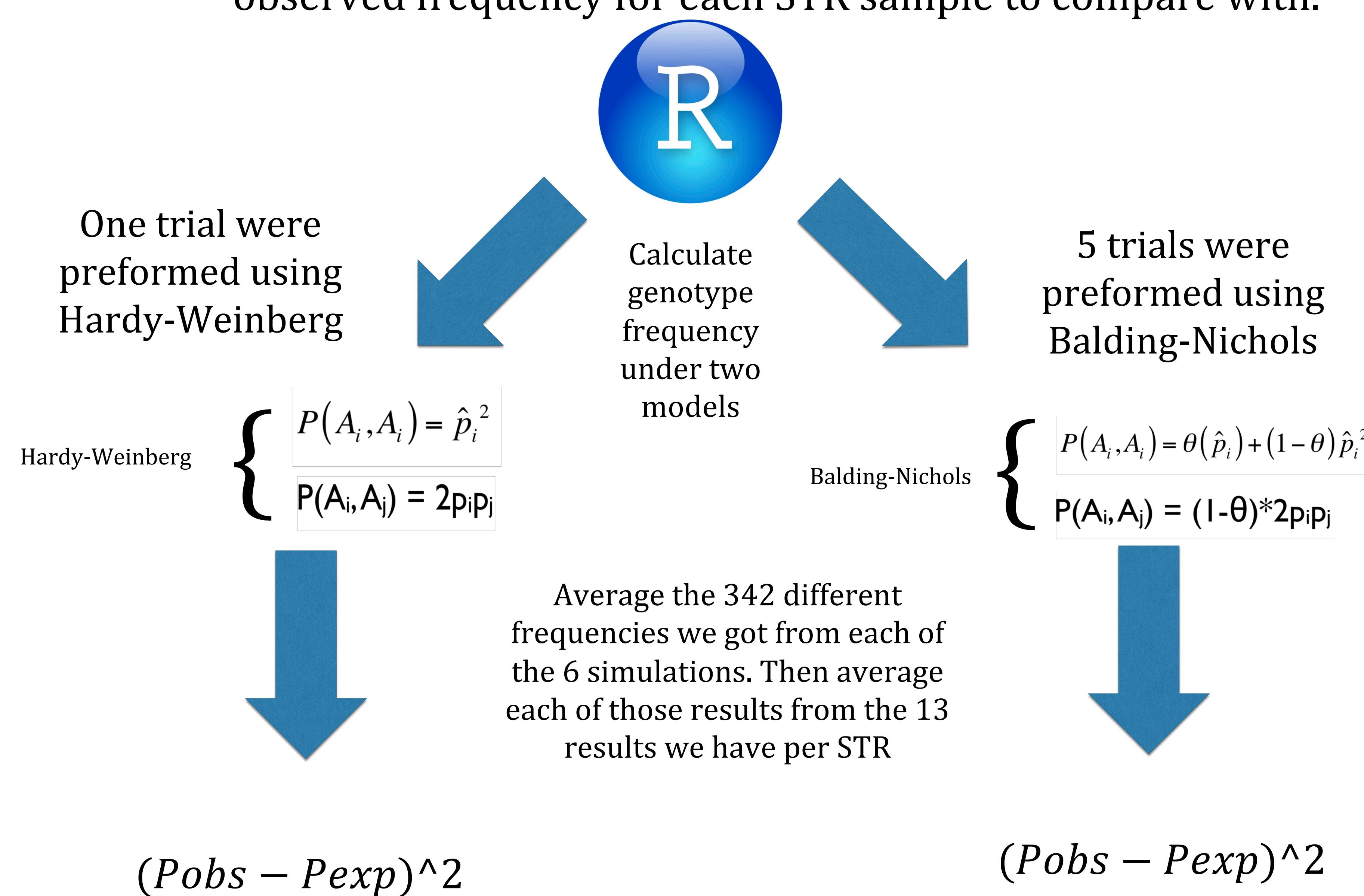
With the use of DNA-Database wide genotype matching, there comes the responsibility in making sure we are making proper genotype probability estimates. We need to make sure that when we are calculating the probability of a match inside a DNA database, we are as accurate as we can be for the sample matched. This is why we need to look at the basic assumptions we make when we calculating these probabilities, like using Hardy-Weinberg for population structure. The Balding-Nichols model is a population structure that includes a factor of correlation of alleles within in a sub population. This change in assumption is important both for how these numbers should be presented in a courtroom, but how they can predict a more accurate genotype frequency in forensic searching.

Question

When calculating genotype frequencies is it more accurate to calculate them under the Balding-Nichols model compared to what is assumed under Hardy-Weinberg.

Method

I used the NIST database looking at allele frequencies to perform my test, I decided only to look inside African Americans population. The database had samples from 342 individual African Americans genotype looking at the allele frequency from all 13 STR samples. We were given in the database the observed frequency for each STR sample to compare with.



References/Acknowledgment

Rori Rohlfs and Vitor Aguiar. "Fitting the Balding-Nichols model to forensic databases", Forensic Science International: Genetics, November 2018

I would like to thank Dr. Rori Rohlfs for teaching an excellent class this semester about forensic genetics and the mathematical applications in this field. I would also like to thank my colleagues Cynthia Perez, Emily Sampeiro, and John Bungay for their help and guidance in this project.

Results

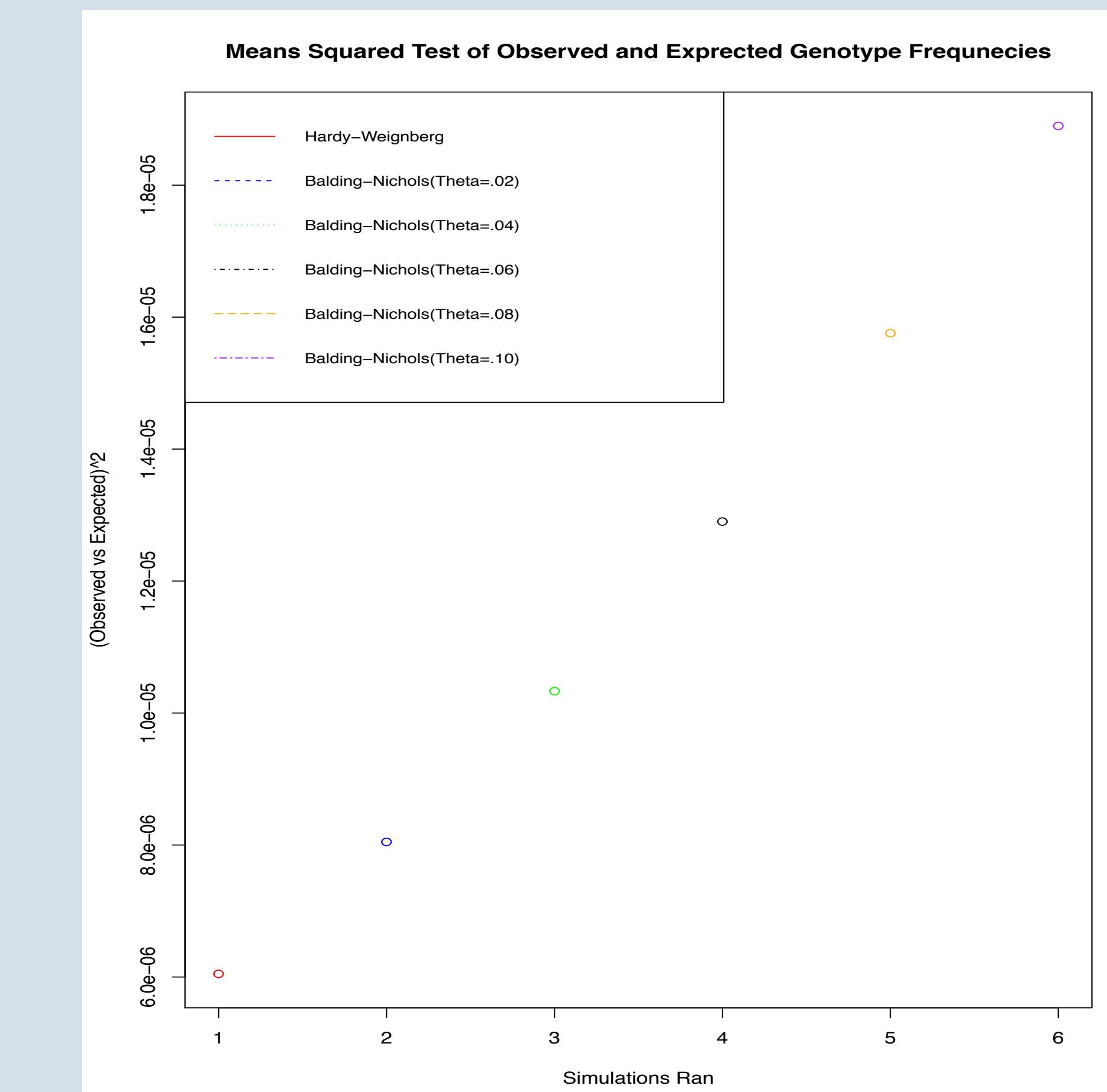


Fig. 1: This graph represents the means squared test result between the 6 simulations ran. This graph shows that as we increase theta the further the distance we are from the observed frequency

Discussion

- We live in a justice system where we are innocent until proven guilty, so we want overestimations to be presented in court.
- The Balding-Nichols model creates an overestimation of a genotype match probability
- Further direction would be to see if theta is effected by ethnicity, or to find a theta that can yield the best result for accurately calculating genotype frequency

Errors in forensics

