Statistica Sinica **35** (2025), 1-27
doi:https://doi.org/10.5705/ss.202023.0128

UNIFICATION OF RARE AND WEAK MULTIPLE TESTING MODELS USING MODERATE DEVIATIONS ANALYSIS AND LOG-CHISQUARED P-VALUES

Alon Kipnis

Reichman University

Abstract: Rare and Weak models for multiple hypothesis testing assume that only a small proportion of the tested hypotheses concern non-null effects and the individual effects are only moderately large, so they generally do not stand out individually, for example in a Bonferroni analysis. Such models have been studied in quite a few settings, for example in some cases studies focused on an underlying Gaussian means model for the hypotheses being tested; in some others, Poisson and Binomial. Such seemingly different models have the following common structure. Summarizing the evidence of individual tests by the negative logarithm of its P-value, the model is asymptotically equivalent to a situation in which most negative log P-values have a standard exponential distribution but a small fraction might have an alternative distribution which is approximately noncentral chisquared on one degree of freedom. This log-chisquared approximation is different from the log-normal approximation of Bahadur. The latter is unsuitable for analyzing Rare and Weak multiple-testing models.

We characterize the asymptotic performance of global tests combining asymptotic log-chisquared P-values in terms of the chisquared mixture parameters: the scaling parameter controlling heteroscedasticity, the non-centrality parameter, and the parameter controlling the rarity of individual non-null effects. In a phase space involving the last two parameters, we derive a region where all tests are asymptotically powerless. Outside of this region, the Berk-Jones and the Higher Criticism tests have maximal power. Inference techniques based on the minimal P-value, false-discovery rate controlling, and Fisher's combination test have sub-optimal asymptotic phase diagrams. Our analysis yields the asymptotic power of global testing in various new rare and weak models, including two-sample heteroscedastic normal mixtures and binomial experiments with perturbed probabilities of success.

Key words and phrases: Heterogeneous mixture, higher criticism, multiple testing, P-values, sparse mixture.

1. Introduction

1.1. Motivation

Consider a multiple hypothesis testing situation, each test involves a different feature of the data where different features are independent. We are interested

*Corresponding author. E-mail: alon.kipnis@runi.ac.il