

Estimation and applications of effect size distributions for large collections of association estimates

Kerby A. Shedden
Department of Statistics
University of Michigan

Abstract: We have been working on the problem of estimating the distribution of effect sizes when a large number of associations are assessed. This is now a familiar setting, with concrete examples being genetic association studies and biomarker screening studies. Technically, the problem involves a deconvolution of distribution functions, which is poorly-posed and hence estimates are somewhat crude. Nevertheless, there are several interesting potential applications: (1) The empirical assessment of effect sparsity, including insights into the structure of large correlation matrices; (2) Estimating the number of associations of a given size within each of several related subpopulations (e.g. disease subtypes), and more interestingly, estimating the proportion of such associations that are shared in pairs of subpopulations; (3) Efficiently identifying associations when related qualitative and quantitative traits are available (e.g. a dichotomous diagnosis and a quantitative symptom score). I will also discuss some connections between this work and FDR, meta-analysis, and related empirical Bayes strategies.