Estimating Distributions in Random Effects Models: Identifiability and Efficiency

Jeffrey D. Hart
Texas A&M University

Suppose observations X_{ij}, $i = 1, \ldots, p$, $j = 1, \ldots, n$, are generated by the following random effects model:

$$X_{ij} = \mu_i + \sigma_i \epsilon_{ij}, \quad i = 1, \ldots, p, \ j = 1, \ldots, n,$$

where the random pairs $(\mu_1, \sigma_1), \ldots, (\mu_p, \sigma_p)$ are i.i.d. as G, the ϵ_{ij}s are i.i.d. as F, and the (μ_i, σ_i) pairs are all independent of the ϵ_{ij}s. This is a common model in microarray analyses and other applications. The scenario of interest is where n is very small in relationship to p. In an asymptotic analysis, this would translate into letting p tend to ∞ with n fixed.

Of interest is obtaining nonparametric estimators of F and G that are consistent (or even efficient) as $p \to \infty$ with n fixed. A little thought convinces one that this is indeed an inverse problem, since the information about F and G in the observed data is rather indirect. A number of interesting questions arise, the most important of which concern identifiability of the model and estimation efficiency. Answers to these questions beg for collaboration between mathematicians and statisticians. The identifiability issue is mathematical, and can be resolved by determining conditions under which certain integral equations have unique solutions. Estimation efficiency is more of a statistical issue. There are at least two broad classes of estimators of F and G that can be applied, and the relative efficiency of the two seems to be a wide open question.

The talk will review some existing results concerning the model defined above, and pose some apparently open problems. More general models will be discussed briefly as well.